Bayesian Networks in Neuroscience

Concha Bielza, Pedro Larrañaga

Computational Intelligence Group
Artificial Intelligence Department
Technical University of Madrid, Spain

PGM'12
Granada, September 19, 2012

Outline

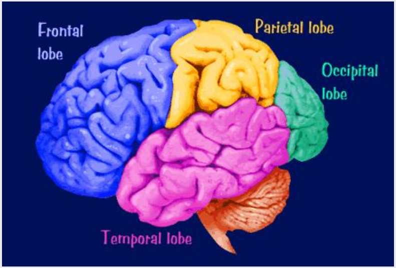
Outline

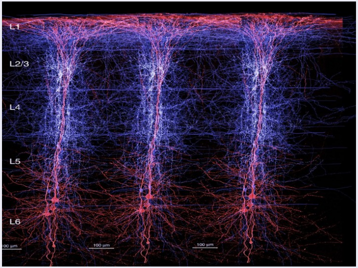
Cajal Blue Brain Project

- At the end of 2008, **Universidad Politécnica de Madrid (UPM)** and **Instituto Cajal (IC)** from the Spanish Research Council, until 2018
- UPM: data analysis, optimization, image analysis and visualization
- IC: morphology and function of neuronal cells

The human brain

Brain lobes and layers

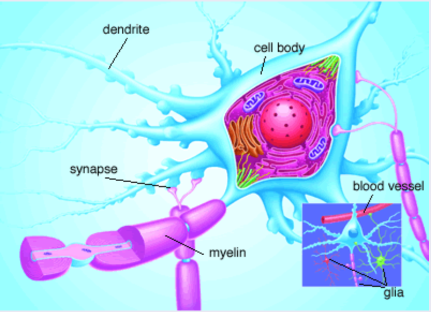




- Weight = 1.3kg, width = 140mm, length = 167mm, height = 93mm

The human brain

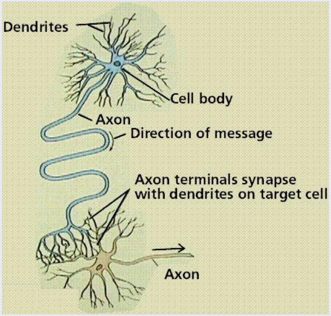
Brain at microscopic level



- Composed of neurons, blood vessels, glial cells
- Neuron is the basic structural and functional unit of the nervous system –neuron doctrine– (S. Ramón y Cajal, late 19th century)
- Just 4 microns thick → could fit 30,000 neurons on the head of a pin
- ∼100,000 million neurons (more than known stars in the universe)

The neuron

3 parts of a neuron: dendrites, soma and axon

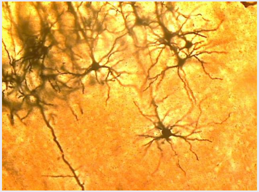


- Axons fill most of the space in the brain $\rightarrow$ $>$150,000 km in the human brain!!
- Each neuron connected to 1,000 neighboring neurons
- 10,000 synaptic connections each

Observing the neurons

Optical (or light) microscope. Stain the tissue

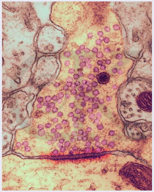
Magnify image up to 2000 times



Golgi's method (1873)

Modern electron microscope

Magnify image up to 2 million times

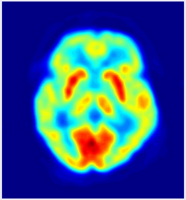


3D from multiple 2D images

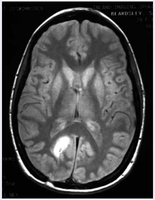
“Visualizing” mental activities from brain images

Electrical activity directly or indirectly

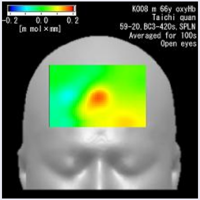
Electroencephalography (EEG)



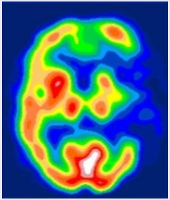
Positron Emission Tomography (PET)



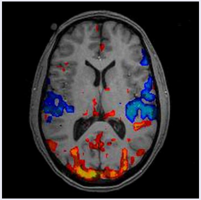
Magnetic Res. Imaging (MRI)



Functional NIR Spectroscopy (fNIRS)



Single Photon Emission Computed Tomography (SPECT)



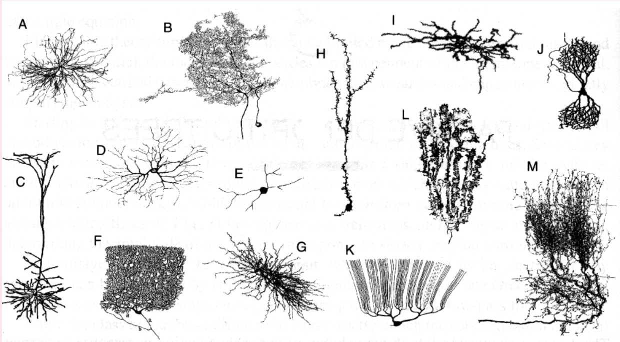
Functional MRI (fMRI)

Outline

Computer simulation of dendritic morphology

Dendritic morphology

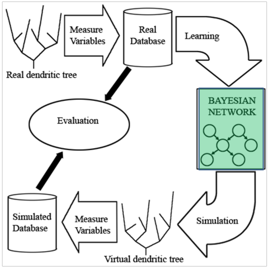
- Tree shapes → interconnectivity and functional roles of neurons
- Their normal function, in neurological diseases, under the effects of some drugs



- Rough groups based on prominent geometrical features. No 2 neurons with the same morphology → but branching patterns

⇒ Anatomical characterization is statistical in nature

Our proposal: advantages



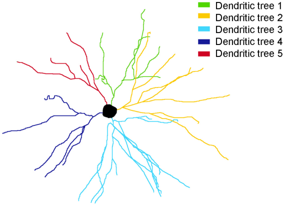
...with Bayesian networks

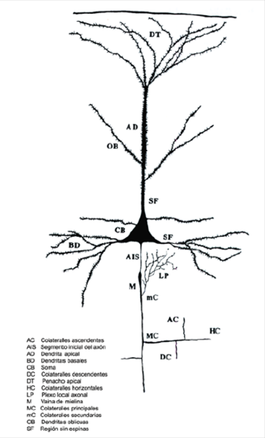
- (In)dependences between morphological properties automatically found from real data (vs. *prior conditional relationships ad hoc*)
- Model the joint probability distribution of all variables (vs. $\leq$ *trivariate* and *standard distributions*)
- Reliable evaluation: statistical tests to compare original vs. simulated distributions, both uni and multivariate (vs. *on new 1D pars and visual inspection*)

López-Cruz, Bielza, Larrañaga, Benavides-Piccione, DeFelipe (2011). Models and simulation of 3D neuronal dendritic trees using Bayesian networks, ***Neuroinformatics***, 9, 347-369

Data: pyramidal neurons

- 3D reconstructions of 90 pyramidal neurons from the mouse neocortex, traced with *Neurolucida* package
- Layer III of different cortical regions: M2, S2, V2L/TeA
  ⇒ 3 databases
- Each basal arbor with 6 (average) main trunks –dendritic trees–, each made up of several dendrites





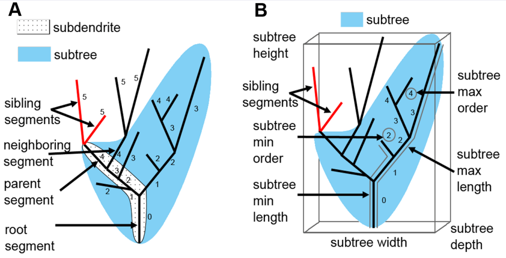
Cortex region	Database	# dendr. trees
Motor	M2	104
Somatosensory	S2	103
Lateral visual and association temporal	V2L/TeA	156

Publicly available at `http://neuromorpho.org` as part of DeFelipe's archive (same lab)

Features

Morphological parameters

- For each pair of sibling segments (line between two branch points), measure 41 variables
- Widely used and also new, to capture context influence and neuritic competition
- Construction variables: define the morphology of a segment (segment length, orientation, bifurcation). Sampled by the model to incrementally construct trees
- Evidence variables: measure the part of the tree previous to a pair of sibling segments (subtree and subdendrite involved). Measured during the simulation, used as information to sample construction variables



List of variables

No.	Type	Variable	No.	Type	Variable
1	E	subtree degree (no. endings)	22	E	neighbor distance
2	E	subtree no. bifurcations (no. nodes)	23	E	neighbor inclination
3	E	subtree total length	24	E	neighbor azimuth
4	E	subtree width	25	E	neighbor extension
5	E	subtree height	26	E	neighbor angle
6	E	subtree depth	27	E	parent segment length
7	E	subtree box volume	28	E	parent segment inclination
8	E	subtree max distance between nodes	29	E	parent segment azimuth
9	E	subtree max distance to soma	30	E	root segment length
10	E	subtree max length	31	E	root segment inclination
11	E	subtree min length	32	E	root segment azimuth
12	E	subtree max order	33	E	segment centrifugal order
13	E	subtree min order	34	C	left segment length
14	E	subdendrite length	35	C	left segment inclination
15	E	subdendrite width	36	C	left segment azimuth
16	E	subdendrite height	37	C	left segment bifurcates
17	E	subdendrite depth	38	C	right segment length
18	E	subdendrite box volume	39	C	right segment inclination
19	E	subdendrite distance to soma	40	C	right segment azimuth
20	E	subdendrite inclination	41	C	right segment bifurcates
21	E	subdendrite azimuth			

Variables **discretized** (2-3 values) trying to preserve empirical distributions

Bayesian network learning

Overview of the learning

- Learn and use a BN for each part of the dendritic tree, to allow specific relationships at each part

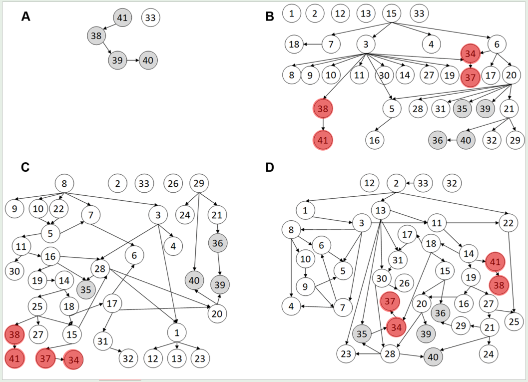
- $P(X_1, ..., X_{41}) = \prod_{i=1}^{n=41} P(X_i|\Pi_i)$ $\Pi_i =$ parents of X_i in the graph
- Learn the structure via K2 algorithm
  - Ordering between nodes *(evidence vars before construction vars)*
  - Fix an upper bound on the max number of parents for any node *(=3)*
- Learn the parameters (probabilities) via MLE

$$P(X_i = x_i|\Pi_i = \pi_i) = \frac{freq(X_i = x_i, \Pi_i = \pi_i)}{freq(\Pi_i = \pi_i)}$$

Bayesian networks learnt

For M2 database

- A, B, C, D $\rightarrow$ root segments, order 1, order 2, $>$ 2 order, resp. Shaded = construction variables

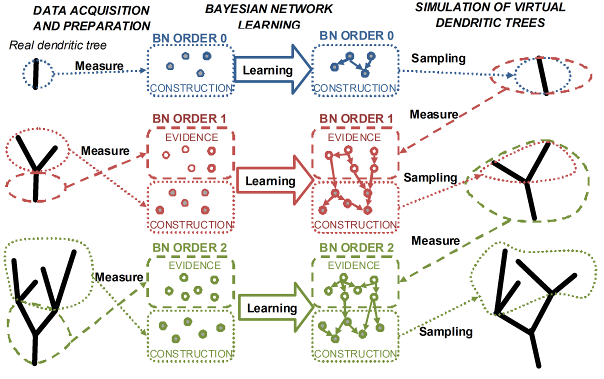


- Found relationships conform to biological knowledge, e.g.
  Segment length (34, 38) and bifurcation (37, 41) occurrence $\rightarrow$ more bifurcations close to the soma and shorter segments, whereas segments that do not branch spread away from the soma

Simulation of virtual dendritic trees

Procedure (breadth-first way)

1. Generate a root segment
2. Measure evidence variables from the dendritic tree built so far
3. Sample construction variable values from the Bayesian network
4. If a segment bifurcates, consider that the dendrite is still incomplete and go to 2. Else, the dendrite has ended



Outline

Outline

Classifying and naming neurons

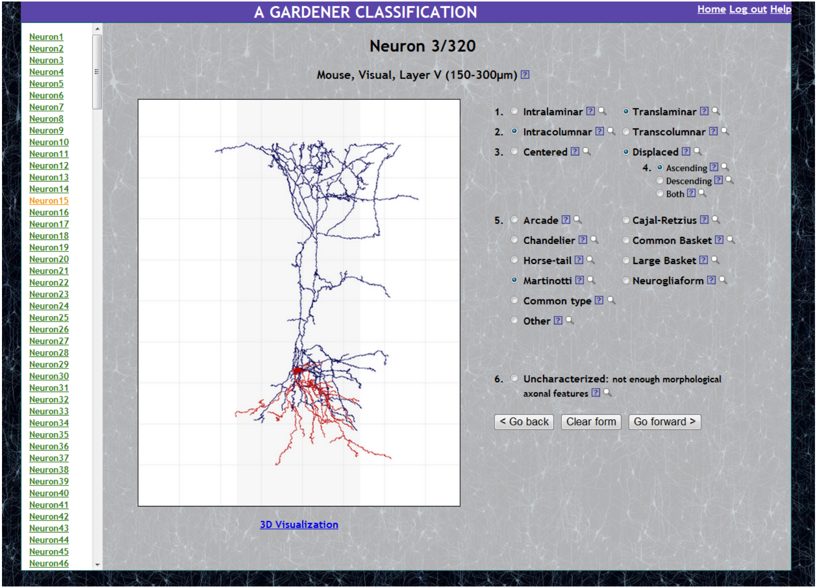
- An accepted catalog of neuron types and names, a debate for over a hundred years since S. Ramón y Cajal
- Amount of data has grown rapidly; better staining methods ⇒ harder classification
- Need of a consistent terminology for an effective communication and data sharing [Petilla Terminology, Ascoli et al. (2008)]
    - Agreement: pyramidal neuron, non-pyramidal, interneuron, chandelier (clear morphological attributes)
    - Disagreement: double bouquet, basket, Martinotti...
    - Virtually every neuroanatomist has his own classification scheme and neuron terms

A ‘gardener’ classification of neurons

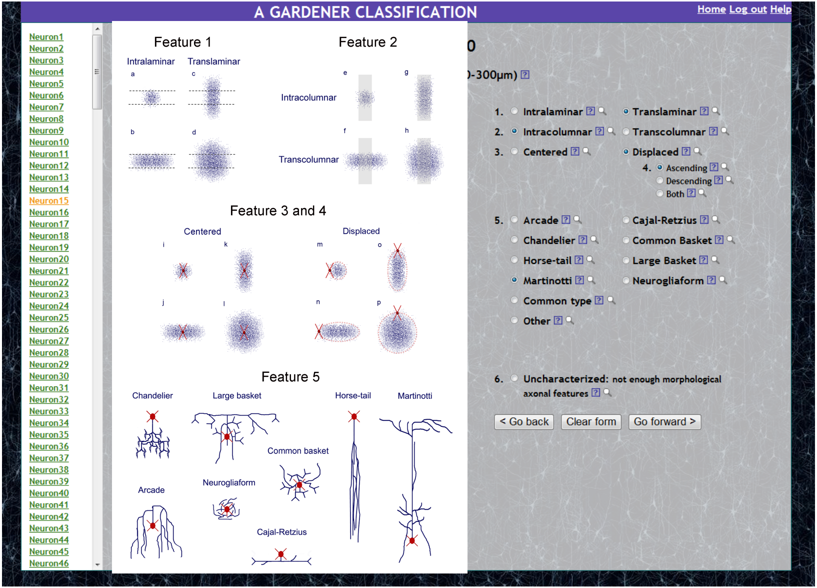
- A ‘gardener’ approach (not a botanist), coarser and practical
- Towards a consensus in naming GABAergic cortical interneurons
  - 10-30% of the total neuron population and main component of inhibitory cortical circuits
  - Located in all cortical layers and with a great variety of morphological, biochemical, and physiological characteristics
- Goal: a community-based strategy for defining a morphological taxonomy, establishing a list of terms to be used by all researchers to distinguish neuronal morphologies

DeFelipe, López-Cruz, Benavides-Piccione, Bielza, Larrañaga, *et al.* (2012). Classification and nomenclature of cortical GABAergic interneurons, ***Nature Reviews Neuroscience***, accepted

Collecting the data: 320 interneurons, 42 experts



Collecting the data: 320 interneurons, 42 experts



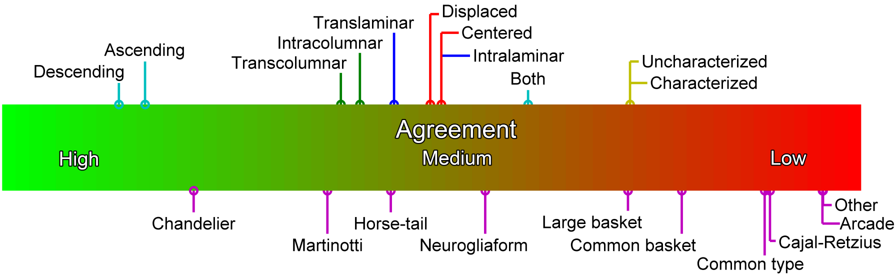
Data

Neuron	Feature 1: E_1	...	E_{42}
1	1	...	0
2	0	...	0
...		...	
320	0	...	0

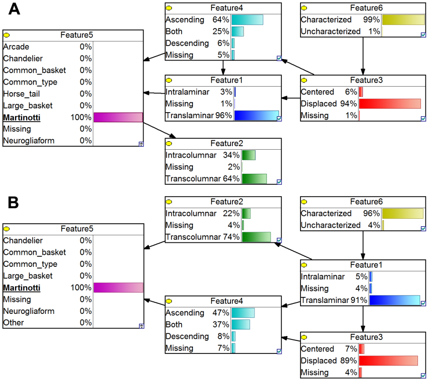
Data

	Feature 1				Feature 5			Feature 6		
Neuron	E_1	...	E_{42}		E_1	...	E_{42}	E_1	...	E_{42}
1	1	...	0		5	...	8	0	...	1
2	0	...	0	....	5	...	10	1	...	1
...		...				...			...	
320	0	...	0		1	...	4	1	...	0

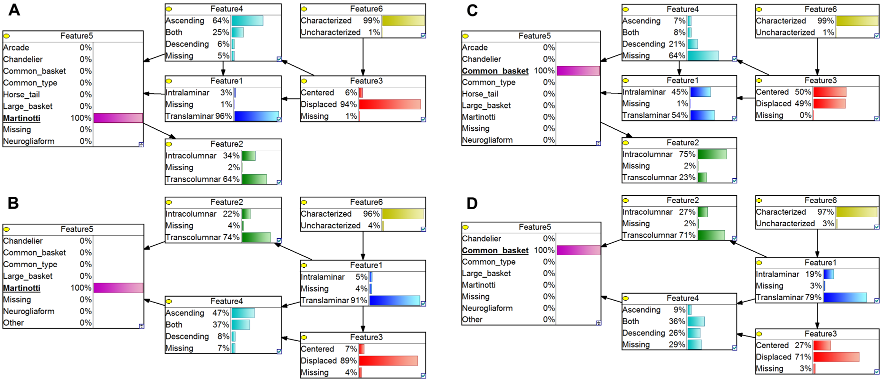
Inter-expert agreement



A Bayesian network learnt for each expert

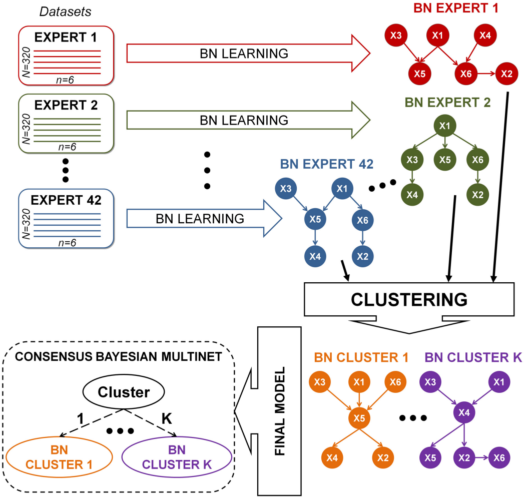


A Bayesian network learnt for each expert



Outline

Inducing a consensus Bayesian multinet from a set of expert opinions



López-Cruz, Larrañaga, DeFelipe, Bielza (2012). Bayesian network modeling of the consensus between experts: An application to neuron classification, ***International J. of Approximate Reasoning***, submitted

Clustering of BNs encoding similar expert opinions

1. Compute the JPD encoded by each BN

2. Cluster the JPDs

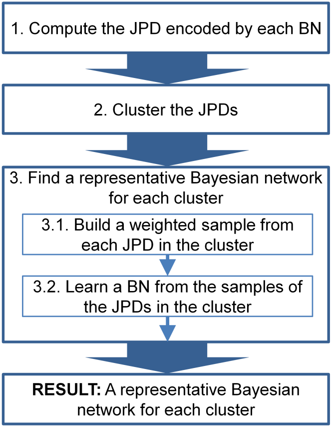
3. Find a representative Bayesian network for each cluster

3.1. Build a weighted sample from each JPD in the cluster

3.2. Learn a BN from the samples of the JPDs in the cluster

RESULT: A representative Bayesian network for each cluster

Clustering of BNs encoding similar expert opinions



Steps 1 and 2

- Dataset with 42 JPDs $\times$ 121 values
- K-means algorithm ($K = 6$)
- Jensen-Shanon divergence as dissimilarity measure for JPDs

$$d_{JS}(\mathbf{p}_1, \mathbf{p}_2) = 0.5\,(KL(\mathbf{p}_1||\mathbf{m}) + KL(\mathbf{p}_2||\mathbf{m}))$$

where $\mathbf{m} = 0.5(\mathbf{p}_1 + \mathbf{p}_2)$

- Compute the cluster center $\bar{\mathbf{p}}_k$ from a set $\{\mathbf{p}_1, \ldots, \mathbf{p}_{N_k}\}$ in cluster k
LOGARITHMIC COMBINATION POOL:

$$\bar{p}_{jLogOp} = \frac{\prod_{i=1}^{N_k} p_{ij}^{\omega_i}}{\sum_{v=1}^{121} \prod_{i=1}^{N_k} p_{iv}^{\omega_i}}$$

with $\omega_i = 1/N_k$

Clustering of BNs encoding similar expert opinions

1. Compute the JPD encoded by each BN

2. Cluster the JPDs

3. Find a representative Bayesian network for each cluster
   - 3.1. Build a weighted sample from each JPD in the cluster
   - 3.2. Learn a BN from the samples of the JPDs in the cluster

RESULT: A representative Bayesian network for each cluster

Step 3

- For each cluster, sample from its JPDs. Draw $\mu_i \times M$ observations from each $\mathbf{p}_i$ in cluster k, where
$$\mu_i = \frac{1 - d_{JS}(\mathbf{p}_i, \overline{\mathbf{p}}_k)}{\sum_{j=1}^{N_k} \left(1 - d_{JS}(\mathbf{p}_j, \overline{\mathbf{p}}_k)\right)}$$
(degree of membership for $\mathbf{p}_i$ to cluster k)
- Learn a (representative) BN from the sample of size M

Cluster labeling (with marginals)

Cluster	# experts
1	3
2	15
3	4
4	12
5	7
6	1

Cluster labeling (with marginals)

Cluster	# experts
1	3
2	15
3	4
4	12
5	7
6	1

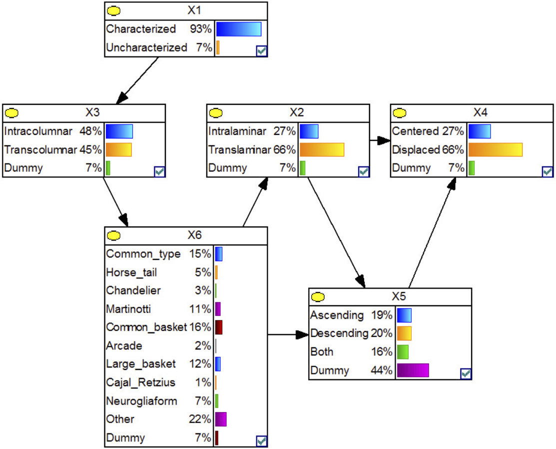
Cluster 4

Coarse classification scheme. High P to *Common type*

Cluster labeling (with marginals)

Cluster	# experts
1	3
2	15
3	4
4	12
5	7
6	1

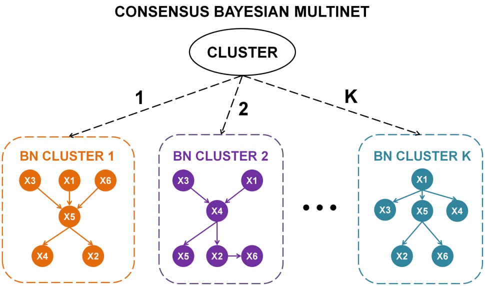
Cluster 5



Detailed classification scheme, distinguishing between *Common type*, *Common basket* and *Large basket*. Found the nomenclature incomplete (high P to *Other*)

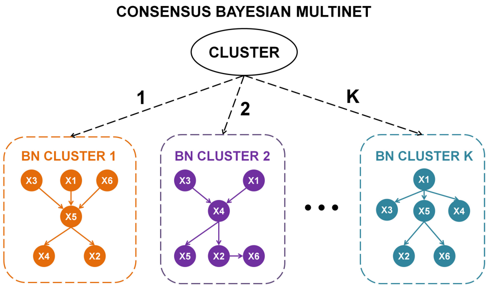
Final consensus Bayesian multinet representing all the experts

- Finite mixture of Bayesian networks: $P(\mathbf{X} = \mathbf{x}) = \sum_{k=1}^{K} \pi_k P_k(\mathbf{X} = \mathbf{x} | C = k)$ with $\pi_k = \frac{N_k}{42}$, P_k=representative BN



Final consensus Bayesian multinet representing all the experts

- Finite mixture of Bayesian networks: $P(\mathbf{X}=\mathbf{x}) = \sum_{k=1}^{K} \pi_k P_k(\mathbf{X}=\mathbf{x}|C=k)$ with $\pi_k = \frac{N_k}{42}$, P_k=representative BN



- Set evidence in X_6 to infer agreed definitions for neuron types:
  - *Martinotti*: Translaminar ($= .93$), Displaced ($= .88$), Ascending ($= .64$)
  - *Common type*: Translaminar ($= .71$)
- Etc.

Outline

Data

	Feature 1				Feature 5			Feature 6		
Neuron	E_1	···	E_{42}		E_1	···	E_{42}	E_1	···	E_{42}
1	1	···	0		5	···	8	0	···	1
2	0	···	0	....	5	···	10	1	···	1
···		···				···			···	
320	0	···	0		1	···	4	1	···	0

Data

Neuron
1
2
...
320

Feature 1

E_1	...	E_{42}
1	...	0
0	...	0
	...	
0	...	0

....

Feature 5

E_1	...	E_{42}
5	...	8
5	...	10
	...	
1	...	4

Feature 6

E_1	...	E_{42}
0	...	1
1	...	1
	...	
1	...	0

128 Axon, 86 Dendrite, 10 Soma; ≥ 26 votes

Neuron	X_1	X_2	...	X_{224}	C
1	8.4	7.1	...	2.1	Martinotti (27)
2	2.0	5.4	...	4.0	?
3	7.3	2.0	...	1.1	?
...		....			...
240	3.7	2.8	...	5.2	Basket (32)

Data

Feature 1

Neuron	E_1	···	E_{42}
1	1	···	0
2	0	···	0
···		···	
320	0	···	0

....

Feature 5

E_1	···	E_{42}
5	···	8
5	···	10
	···	
1	···	4

Feature 6

E_1	···	E_{42}
0	···	1
1	···	1
	···	
1	···	0

128 Axon, 86 Dendrite, 10 Soma | ≥ 26 votes

Neuron	X_1	X_2	···	X_{224}	C
1	8.4	7.1	···	2.1	Martinotti (27)
2	2.0	5.4	···	4.0	?
3	7.3	2.0	···	1.1	?
···		....			···
240	3.7	2.8	···	5.2	Basket (32)

- Unlabeled instances to be classified not only according to the known labels but also discovering new unknown clusters $\rightarrow$ **Semi-supervised clustering**

Data

Neuron	Feature 1: E_1	...	E_{42}
1	1	...	0
2	0	...	0
...		...	
320	0	...	0

....

Feature 5: E_1	...	E_{42}	Feature 6: E_1	...	E_{42}
5	...	8	0	...	1
5	...	10	1	...	1
	...			...	
1	...	4	1	...	0

128 Axon, 86 Dendrite, 10 Soma; $\geq$ 26 votes

Neuron	X_1	X_2	...	X_{224}	C
1	8.4	7.1	...	2.1	Martinotti (27)
2	2.0	5.4	...	4.0	?
3	7.3	2.0	...	1.1	?
...		....			...
240	3.7	2.8	...	5.2	Basket (32)

- Unlabeled instances to be classified not only according to the known labels but also discovering new unknown clusters $\rightarrow$ **Semi-supervised clustering**
- Localized FSS, each cluster described by a different FS $\rightarrow$ **Subspace** clustering

Semi-supervised subspace probabilistic clustering

- Probabilistic clustering → Estimate a (finite mixture) model
- Latent variables set → $\mathcal{Z} = \mathcal{Z}^{\mathcal{L}} \cup \mathcal{Z}^{\mathcal{U}} = \{\mathbf{z}_1, \ldots, \mathbf{z}_L\} \cup \{\mathbf{z}_{L+1}, \ldots, \mathbf{z}_N\}$
- $\mathbf{z}_i = (0_1, 0_2, ..., 1_m, ..., 0_K)$ if instance i belongs to component m; $p(z_{im} = 1) = \pi_m$

- $p(\mathbf{x}_i \mid \Theta) = \sum_{m=1}^{K} \pi_m p(\mathbf{x}_i \mid \boldsymbol{\theta}_m)$ Density
- $\log L(\Theta \mid \mathcal{X}, \mathcal{Z}) = \sum_{i=1}^{N} \sum_{m=1}^{K} z_{im}(\log \pi_m + \log p(\mathbf{x}_i \mid \theta_m))$ Complete-data log-lik
- $\mathcal{Q}(\Theta, \Theta^{t-1}) = \mathbb{E}_{\mathcal{Z}|\mathcal{X},\Theta^{t-1}}[\log L(\Theta \mid \mathcal{X}, \mathcal{Z})]$ Its expectation (E-step)
- $\Theta^t = \arg\max_{\Theta} \mathcal{Q}(\Theta, \Theta^{t-1})$ Its max (M-step)

Semi-supervised subspace probabilistic clustering

- Subspaces: More latent vars $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_K\}$, with $\mathbf{v}_m = (0_1, 1_2, ..., 1_j, ..., 0_F)$ if features 2 and j are relevant to component m; $p(v_{mj} = 1) = \rho_{mj}$
- z_{im} indicates instance i's membership of component m; v_{mj} indicates feature j's relevance to component m

- $p(\mathbf{x}_i|\Theta)$
  $= \sum_m \pi_m \prod_j \left(\rho_{mj} p(x_{ij}|\theta_{mj}) + (1-\rho_{mj}) p(x_{ij}|\lambda_{mj})\right)$
  Density (assume c.i. of the features given the component)
- $\log L(\Theta \mid \mathcal{X}, \mathcal{Z}, \mathcal{V})$
- $\mathbb{E}_{\mathcal{Z},\mathcal{V}|\mathcal{X},\Theta^{t-1}}$
- $\Theta^t = \text{arg max}_{\Theta} \mathcal{Q}(\Theta, \Theta^{t-1})$

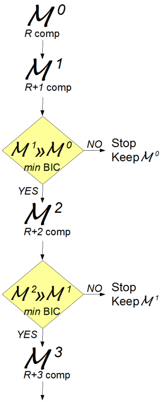
- Semi-supervised: include label info
- Labeled $\mapsto$ known classes $\{1, \ldots, R\}$;
  Unlabeled $\mapsto$ any $\{1, ..., R, \ldots, K\}$
  $\mathbb{E}_{\mathcal{Z},\mathcal{V}|\mathcal{X},\Theta^{t-1}} = \mathbb{E}_{\mathcal{Z}^L,\mathcal{V}|\mathcal{X}^L,\Theta^{t-1}} + \mathbb{E}_{\mathcal{Z}^U,\mathcal{V}|\mathcal{X}^U,\Theta^{t-1}}$

Semi-supervised subspace probabilistic clustering

- Subspaces: More latent vars $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_K\}$, with $\mathbf{v}_m = (0_1, 1_2, ..., 1_j, ..., 0_F)$ if features 2 and j are relevant to component m; $p(v_{mj} = 1) = \rho_{mj}$
- z_{im} indicates instance i's membership of component m; v_{mj} indicates feature j's relevance to component m

- $p(\mathbf{x}_i|\Theta)$
  $= \sum_m \pi_m \prod_j \Big(\rho_{mj} p(x_{ij}|\theta_{mj}) + (1-\rho_{mj}) p(x_{ij}|\lambda_{mj})\Big)$
  Density (assume c.i. of the features given the component)
- $\log L(\Theta \mid \mathcal{X}, \mathcal{Z}, \mathcal{V})$
- $\mathbb{E}_{\mathcal{Z},\mathcal{V}|\mathcal{X},\Theta^{t-1}}$
- $\Theta^t = \arg\max_{\Theta} \mathcal{Q}(\Theta, \Theta^{t-1})$

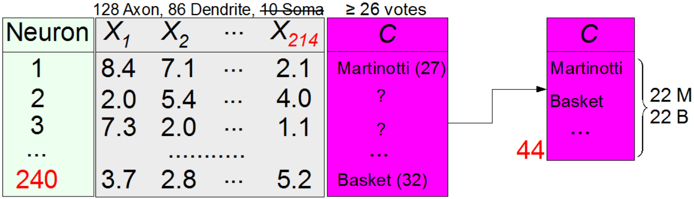
- Semi-supervised: include label info
- Labeled $\mapsto$ known classes $\{1, \ldots, R\}$;
  Unlabeled $\mapsto$ any $\{1, ..., R, \ldots, K\}$
  $\mathbb{E}_{\mathcal{Z},\mathcal{V}|\mathcal{X},\Theta^{t-1}} = \mathbb{E}_{\mathcal{Z}^L,\mathcal{V}|\mathcal{X}^L,\Theta^{t-1}} + \mathbb{E}_{\mathcal{Z}^U,\mathcal{V}|\mathcal{X}^U,\Theta^{t-1}}$



Guerra, Bielza, Robles, Larrañaga (2012). Semi-supervised subspace model-based clustering, ***Data Mining and Knowledge Discovery***, submitted

Martinotti and Basket cells

- Take Martinotti and Basket (common and large basket: distinction not clear)
- Which parts of the neurons (axon, dendrites) are more important for distinguishing them and finding new subtypes

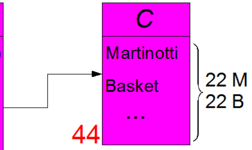


Martinotti and Basket cells

- Take Martinotti and Basket (common and large basket: distinction not clear)
- Which parts of the neurons (axon, dendrites) are more important for distinguishing them and finding new subtypes

128 Axon, 86 Dendrite, ~~10 Soma~~ ≥ 26 votes

Neuron	X_1	X_2	···	X_{214}	C
1	8.4	7.1	···	2.1	Martinotti (27)
2	2.0	5.4	···	4.0	?
3	7.3	2.0	···	1.1	?
···	....				···
240	3.7	2.8	···	5.2	Basket (32)



- Since labeled instances don't change their labels, we look at the unlabeled instances

1. Hiding M: C-column with {B,?}

M M M M M M M
M M M M M M M
M M M M M M M M

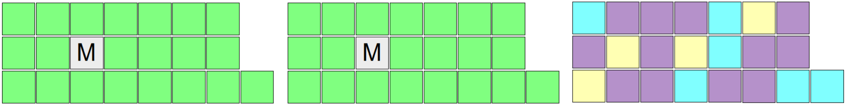
2. Hiding B: C-column with {M,?}

B B B B B B B
B B B B B B B
B B B B B B B B

Martinotti and Basket cells

2. Hiding B

All | Axon | Dendrites

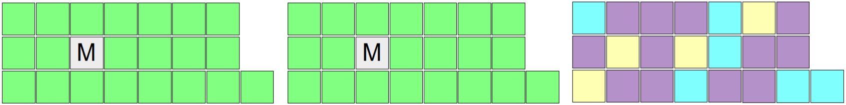


1 new subtype; 1 error | 1 new subtype; 1 error | 3 new subtypes; 0 errors

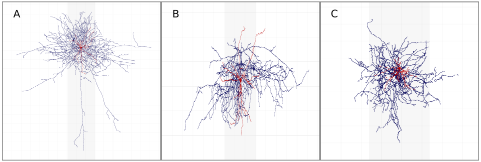
Martinotti and Basket cells

2. Hiding B

All | Axon | Dendrites

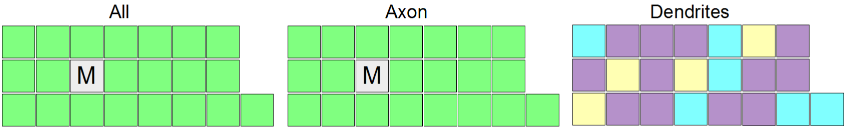


1 new subtype; 1 error | 1 new subtype; 1 error | 3 new subtypes; 0 errors

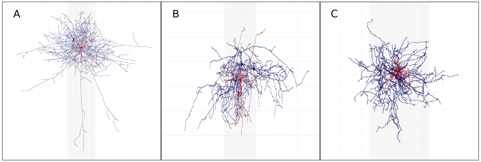


Martinotti and Basket cells

2. Hiding B



1 new subtype; 1 error — 1 new subtype; 1 error — 3 new subtypes; 0 errors



- Axonal features traditionally considered the most important to classify neurons
- However, dendritic features identified new B groups, while the main characteristics of them are related to the axon

⇒ Dendritic characteristics in neurons could be more related to axonal than previously believed

Outline

Data

	Feature 1			...	Feature 5			Feature 6		
Neuron	E_1	...	E_{42}		E_1	...	E_{42}	E_1	...	E_{42}
1	1	...	0		5	...	8	0	...	1
2	0	...	0		5	...	10	1	...	1
...		...				...			...	
320	0	...	0		1	...	4	1	...	0

Data

Neuron	Feature 1 E_1	...	Feature 1 E_{42}	....	Feature 5 E_1	...	Feature 5 E_{42}	Feature 6 E_1	...	Feature 6 E_{42}
1	1	...	0		5	...	8	0	...	1
2	0	...	0		5	...	10	1	...	1
...		...				...			...	
320	0	...	0		1	...	4	1	...	0

Neuron	X_1	X_2	...	X_{2886}	Count vector with votes: *Intralaminar* vs *Translaminar*	Probability distribution: *Intralaminar* vs *Translaminar*
1	8.4	7.1	...	2.1	(38, 4)	(.90,.10)
2	2.0	5.4	...	4.0	(3, 39)	(.07,.93)
3	7.3	2.0	...	1.1	(10,32)	(.24,.76)
...		....			...	...
240	3.7	2.8	...	5.2	(24,18)	(.57,.43)

Data

Neuron	Feature 1 E_1	…	E_{42}	……………………	Feature 5 E_1	…	E_{42}	Feature 6 E_1	…	E_{42}
1	1	…	0		5	…	8	0	…	1
2	0	…	0		5	…	10	1	…	1
…		…				…			…	
320	0	…	0		1	…	4	1	…	0

Neuron	X_1	X_2	…	X_{2886}	Count vector with votes *Intralaminar vs Translaminar*	Probability distribution *Intralaminar vs Translaminar*
1	8.4	7.1	…	2.1	(38, 4)	(.90,.10)
2	2.0	5.4	…	4.0	(3, 39)	(.07,.93)
3	7.3	2.0	…	1.1	(10,32)	(.24,.76)
…			………..		…	…
240	3.7	2.8	…	5.2	(24,18)	(.57,.43)

- A **probability distribution over the class labels** $\mathbf{p}_i = \{p_{ic}\}_{c \in \Omega_C}$ for each instance i

Probabilistic label EM: PLEM algorithm

- Framework for learning Gaussian (mixture) classifiers [Côme et al., 2009]

$$f(\mathbf{x}) = \sum_{c=1}^{K} \pi_c f_{\mathbf{X}|c}(\mathbf{x}; \boldsymbol{\mu}_{\mathbf{X}|c}, \boldsymbol{\Sigma}_{\mathbf{X}|c})$$

where the class information is modeled as probability distributions

Probabilistic label EM: PLEM algorithm

- Framework for learning Gaussian (mixture) classifiers [Côme et al., 2009]

$$f(\mathbf{x}) = \sum_{c=1}^{K} \pi_c f_{\mathbf{X}|c}(\mathbf{x}; \boldsymbol{\mu}_{\mathbf{X}|c}, \boldsymbol{\Sigma}_{\mathbf{X}|c})$$

  where the class information is modeled as probability distributions

- $LL = \ln(p(\Theta|\mathcal{D})) = \sum_{i=1}^{N} \ln \left(\sum_{c=1}^{K} p_{ic} \pi_c f_{\mathbf{X}|c}(\mathbf{x}_i; \boldsymbol{\mu}_{\mathbf{X}|c}, \boldsymbol{\Sigma}_{\mathbf{X}|c}) \right)$ Generalized log-lik

Probabilistic label EM: PLEM algorithm

- Framework for learning Gaussian (mixture) classifiers [Côme et al., 2009]

$$f(\mathbf{x}) = \sum_{c=1}^{K} \pi_c f_{\mathbf{X}|c}(\mathbf{x}; \boldsymbol{\mu}_{\mathbf{X}|c}, \boldsymbol{\Sigma}_{\mathbf{X}|c})$$

  where the class information is modeled as probability distributions

- $LL = \ln(p(\Theta|\mathcal{D})) = \sum_{i=1}^{N} \ln\left(\sum_{c=1}^{K} p_{ic}\pi_c f_{\mathbf{X}|c}(\mathbf{x}_i; \boldsymbol{\mu}_{\mathbf{X}|c}, \boldsymbol{\Sigma}_{\mathbf{X}|c})\right)$ Generalized log-lik

Generalized EM

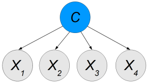
- E-step: $t_{ic}^{(q)} = \dfrac{p_{ic}\pi_c^{(q)} f_{\mathbf{X}|c}(\mathbf{x}_i; \boldsymbol{\mu}_{\mathbf{X}|c}^{(q)}, \boldsymbol{\Sigma}_{\mathbf{X}|c}^{(q)})}{\sum_{c'=1}^{K} p_{ic'}\pi_{c'}^{(q)} f_{\mathbf{X}|c'}(\mathbf{x}_i; \boldsymbol{\mu}_{\mathbf{X}|c'}^{(q)}, \boldsymbol{\Sigma}_{\mathbf{X}|c'}^{(q)})}$. We set $t_{ic}^{(0)} \leftarrow p_{ic}$
- M-step: $\pi_c^{(q+1)} = \frac{1}{N}\sum_{i=1}^{N} t_{ic}^{(q)}$, $\quad \boldsymbol{\mu}_{\mathbf{X}|c}^{(q+1)} = \frac{1}{\sum_{i=1}^{N} t_{ic}^{(q)}} \sum_{i=1}^{N} t_{ic}^{(q)}\mathbf{x}_i$, and
  $\boldsymbol{\Sigma}_{\mathbf{X}|c}^{(q+1)} = \frac{1}{\sum_{i=1}^{N} t_{ic}^{(q)}} \sum_{i=1}^{N} t_{ic}^{(q)}(\mathbf{x}_i - \boldsymbol{\mu}_{\mathbf{X}|c}^{(q+1)})(\mathbf{x}_i - \boldsymbol{\mu}_{\mathbf{X}|c}^{(q+1)})^T$

López-Cruz, Bielza, Larrañaga (2012). Learning conditional linear Gaussian classifiers from class label counts using finite mixture models, ***Journal of Artificial Intelligence Research***, submitted

Learning Bayesian classifiers with PLEM

- NB classifier:

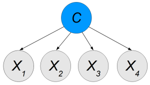
$$f_{\mathbf{X}}(\mathbf{x}) = \sum_{c=1}^{K} \pi_c \prod_{j=1}^{n} f_{X_j|c}(x_j)$$



Learning Bayesian classifiers with PLEM

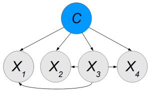
- NB classifier:

$$f_{\mathbf{X}}(\mathbf{x}) = \sum_{c=1}^{K} \pi_c \prod_{j=1}^{n} f_{X_j|c}(x_j)$$



- AODE classifier, averaging the n predictions of:

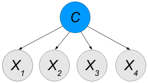
$$f_{\mathbf{X}}(\mathbf{x}) = \sum_{c=1}^{K} \pi_c f_{X_j|c}(x_j) \prod_{k=1, k \neq j}^{n} f_{X_k|X_j,c}(x_k)$$



Learning Bayesian classifiers with PLEM

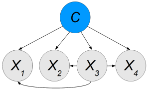
- NB classifier:

$$f_{\mathbf{X}}(\mathbf{x}) = \sum_{c=1}^{K} \pi_c \prod_{j=1}^{n} f_{X_j|c}(x_j)$$



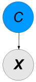
- AODE classifier, averaging the *n* predictions of:

$$f_{\mathbf{X}}(\mathbf{x}) = \sum_{c=1}^{K} \pi_c f_{X_j|c}(x_j) \prod_{k=1,k\neq j}^{n} f_{X_k|X_j,c}(x_k)$$



- Multivariate Gaussian classifier:

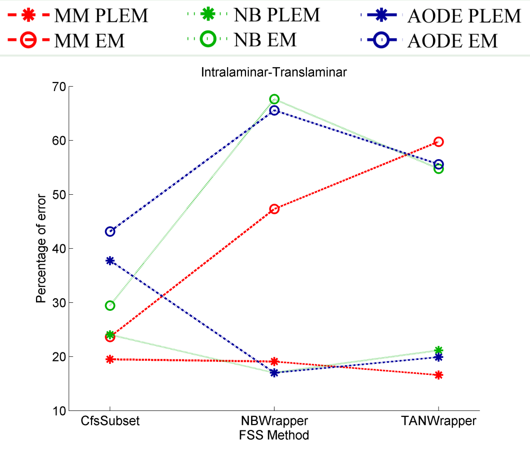
$$f_{\mathbf{X}}(\mathbf{x}) = \sum_{c=1}^{K} \pi_c f_{\mathbf{X}|c}(\mathbf{x})$$



PLEM vs EM

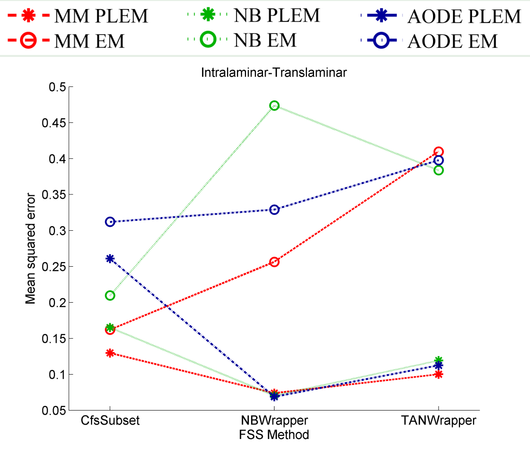
- EM: Set initial posterior probabilities (in E-step) as p_{ic}
- First, 3 FSS methods: CFS, NBWrapper, TANWrapper

Results on classification error: compare Mode in p_i with predicted Mode



PLEM vs EM

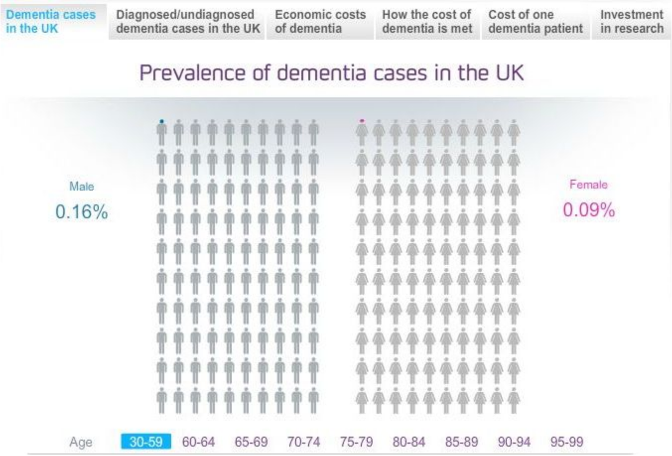
Results on mean squared error: $MSE = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{K}\sum_{c=1}^{K}(p_{ic}-\hat{p}_{ic})^2$



Outline

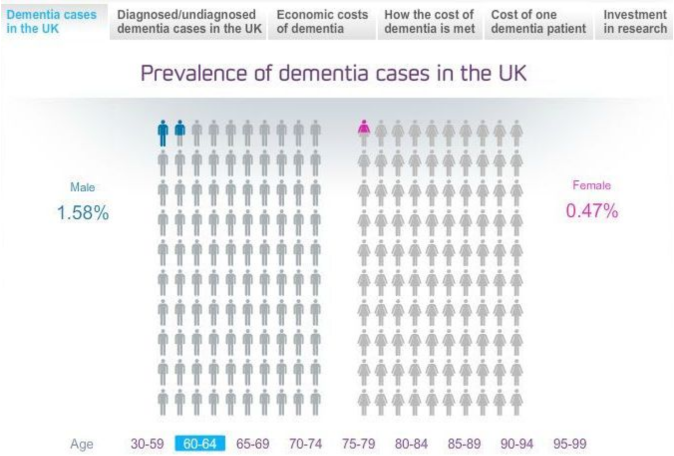
Outline

Dementia: Prevalence, cost and investment in research



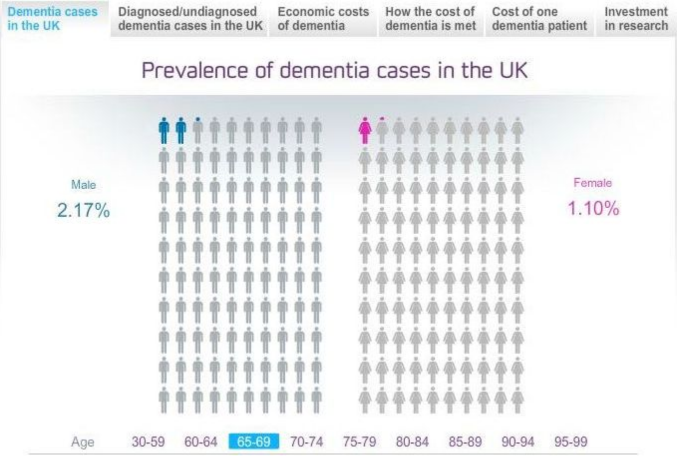
http://www.alzheimersresearchuk.org/dementia-statistics/

Dementia: Prevalence, cost and investment in research



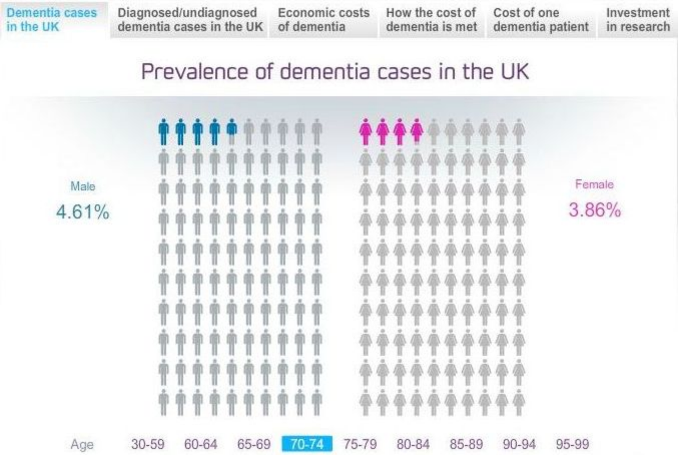
http://www.alzheimersresearchuk.org/dementia-statistics/

Dementia: Prevalence, cost and investment in research



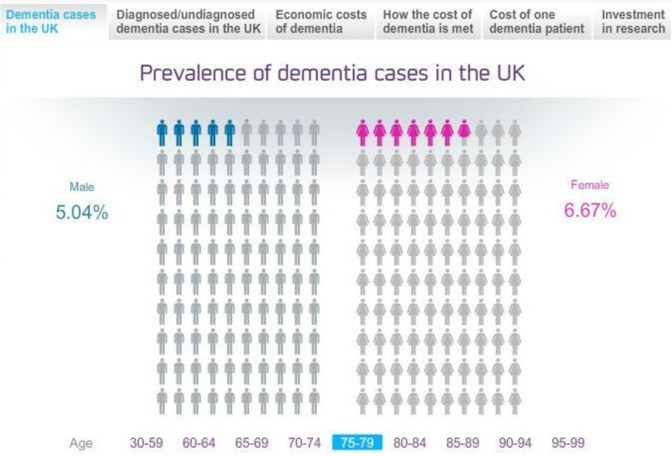
http://www.alzheimersresearchuk.org/dementia-statistics/

Dementia: Prevalence, cost and investment in research



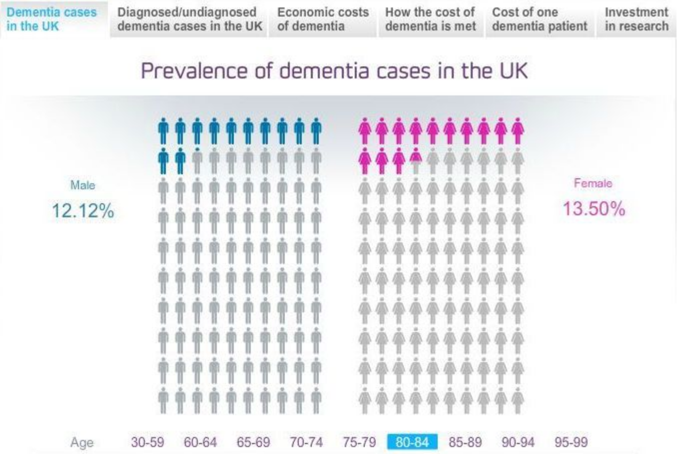
http://www.alzheimersresearchuk.org/dementia-statistics/

Dementia: Prevalence, cost and investment in research



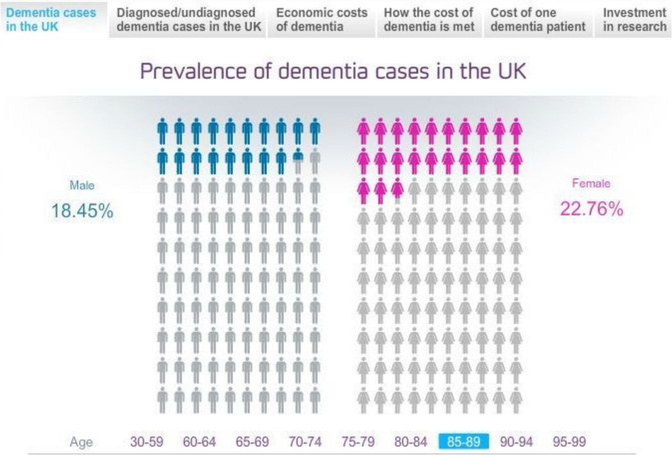
http://www.alzheimersresearchuk.org/dementia-statistics/

Dementia: Prevalence, cost and investment in research



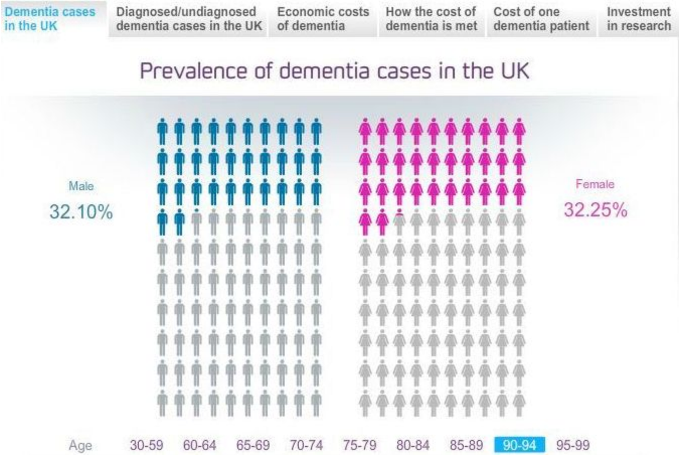
http://www.alzheimersresearchuk.org/dementia-statistics/

Dementia: Prevalence, cost and investment in research



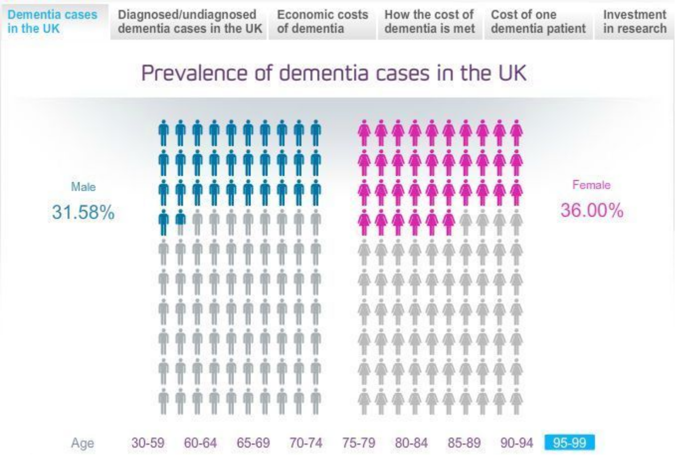
http://www.alzheimersresearchuk.org/dementia-statistics/

Dementia: Prevalence, cost and investment in research



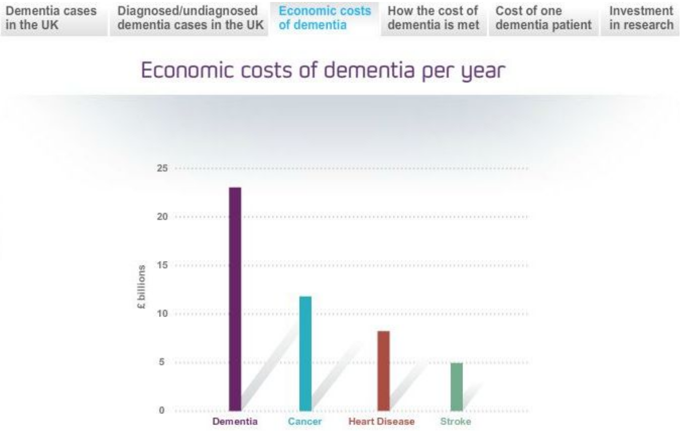
http://www.alzheimersresearchuk.org/dementia-statistics/

Dementia: Prevalence, cost and investment in research



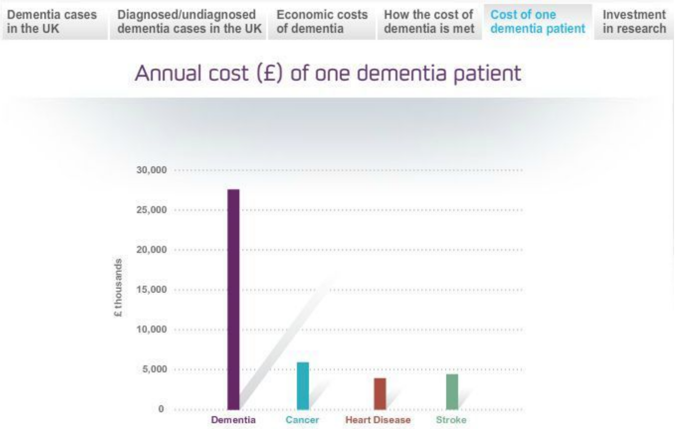
http://www.alzheimersresearchuk.org/dementia-statistics/

Dementia: Prevalence, cost and investment in research



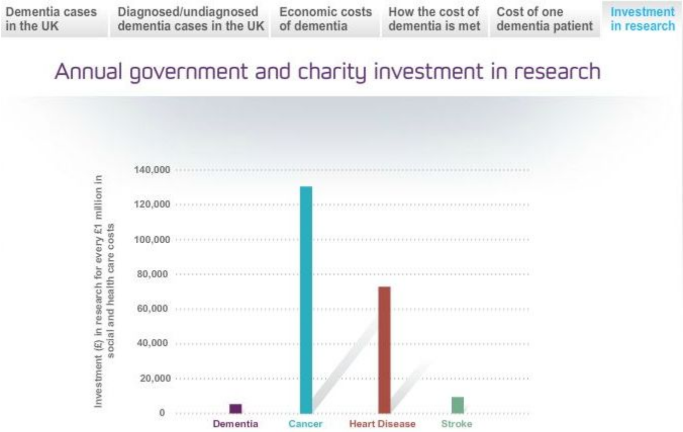
http://www.alzheimersresearchuk.org/dementia-statistics/

Dementia: Prevalence, cost and investment in research



http://www.alzheimersresearchuk.org/dementia-statistics/

Dementia: Prevalence, cost and investment in research



http://www.alzheimersresearchuk.org/dementia-statistics/

Web of Knowledge (Thomson Reuters): “Bayesian network + ...” on Sept 1, 2012

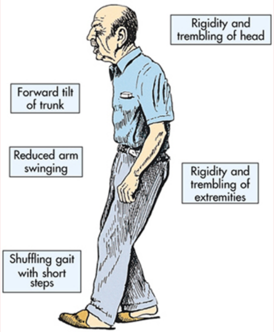
Disease	No. articles	No. citations
Alzheimer	23	426
Parkinson	13	96
Autism	2	8
Schizophrenia	24	116
Multiple sclerosis	6	21

Outline

Supervised classification of dementia development in Parkinson's disease

Dementia development in Parkinson's disease (PD)

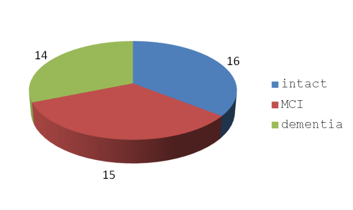
- PD: 1% of the population $>$ 60 years old
- Dementia affects $\sim$40% of PD patients
- Objectives:
  - Discriminate between PD patients cognitively intact, mild cognitive impairment (MCI) and dementia
  - Identify the most predictive neuroanatomic biomarkers (vs previous MRI studies with assumed preselected structure)



Morales, Larrañaga, Bielza, et al. (2012). Predicting dementia development in Parkinson's disease using Bayesian network classifiers, ***Psychiatry Research. Neuroimaging***, accepted

Supervised classification of dementia development in Parkinson's disease

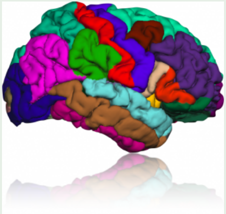
45 Patients



Labels

Hospital Santa Creu i Sant Pau

FreeSurfer: 112 variables



Supervised classification of dementia development in Parkinson's disease

Accuracy results, with 5-fold cross-validation

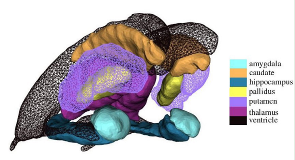
- Kruskal-Wallis non-parametric test with $\alpha = 0.05$

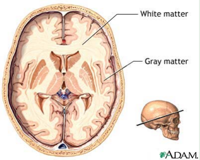
Classifier	intact-dementia	intact-MCI	MCI-dementia	intact-MCI-dementia
Naive Bayes	93.33$\pm$ 9.12	86.66$\pm$13.40	**96.55$\pm$ 7.85**	64.44$\pm$14.48
Selective NB-filter	93.33$\pm$10.66	**89.00$\pm$14.48**	**96.66$\pm$10.33**	70.00$\pm$26.66
Selective NB-CFS	**97.00$\pm$ 6.74**	**90.09$\pm$ 8.40**	**96.55$\pm$ 7.85**	68.88$\pm$16.48
SVM	**96.67$\pm$10.82**	84.10$\pm$15.94	79.31$\pm$13.84	62.22$\pm$18.59

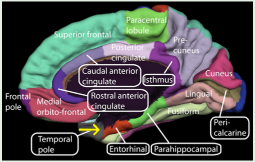
- FSS improved performance in general
- Different relevant variables in each classification problem are automatically identified

Supervised classification of dementia development in Parkinson's disease

Selected features







1. `intact` vs `dementia`: left and right inferior lateral ventricles (+), left white matter (-), left hippocampus (-), right lateral ventricle (-), left cerebellum white matter (-), and right entorhinal (-)
2. `intact` vs `MCI`: brain stem and left hippocampus
3. `MCI` vs `dementia`: left cerebral cortex, left caudate, right inferior lateral ventricle and left entorhinal
4. `intact` vs `MCI` vs `dementia`: left thalamus proper, right inferior lateral ventricle, left caudal anterior cingulate, left entorhinal and left fusiform

Outline

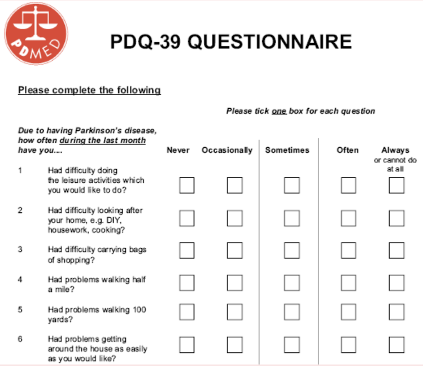
Multi-dimensional classification for EQ-5D health states from PDQ-39 in Parkinson

PDQ-39 and EQ-5D: quality of life instruments to measure the degree of disability

PDQ-39

PDQ-39 captures patient's perception of his illness covering 8 dimensions:

1. Mobility
2. Activities of daily living
3. Emotional well-being
4. Stigma
5. Social support
6. Cognitions
7. Communication
8. Bodily discomfort



PDQ-39 QUESTIONNAIRE

Please complete the following

Please tick one box for each question

Due to having Parkinson's disease, how often during the last month have you....		Never	Occasionally	Sometimes	Often	Always or cannot do at all
1	Had difficulty doing the leisure activities which you would like to do?	☐	☐	☐	☐	☐
2	Had difficulty looking after your home, e.g. DIY, housework, cooking?	☐	☐	☐	☐	☐
3	Had difficulty carrying bags of shopping?	☐	☐	☐	☐	☐
4	Had problems walking half a mile?	☐	☐	☐	☐	☐
5	Had problems walking 100 yards?	☐	☐	☐	☐	☐
6	Had problems getting around the house as easily as you would like?	☐	☐	☐	☐	☐

Multi-dimensional classification for EQ-5D health states from PDQ-39 in Parkinson

EQ-5D

EQ-5D is a generic measure of health for clinical and economic appraisal

Mobility
I have no problems in walking about
I have some problems in walking about
I am confined to bed

Self-care
I have no problems with self-care
I have some problems washing and dressing myself
I am unable to wash and dress myself

Usual activities (eg. work, study, housework, family or leisure activities)
I have no problems with performing my usual activities
I have some problems with performing my usual activities
I am unable to perform my usual activities

Pain/discomfort
I have no pain or discomfort
I have moderate pain or discomfort
I have extreme pain or discomfort

Anxiety/depression
I am not anxious or depressed
I am moderately anxious or depressed
I am extremely anxious or depressed

Multi-dimensional classification for EQ-5D health states from PDQ-39 in Parkinson

Mapping PDQ-39 to EQ-5D

PDQ_1	PDQ_2	...	...	PDQ_{39}	EQ_1	EQ_2	EQ_3	EQ_4	EQ_5
3	1	...	...	3	1	3	3	2	1
2	3	...	...	2	1	1	2	3	2
5	2	...	...	4	1	3	3	1	2
...	...	...	...	...	...	...	...	...	...
4	4	...	...	3	3	1	2	3	2
4	4	...	...	3	3	1	2	3	2
5	5	...	...	4	2	3	2	3	3

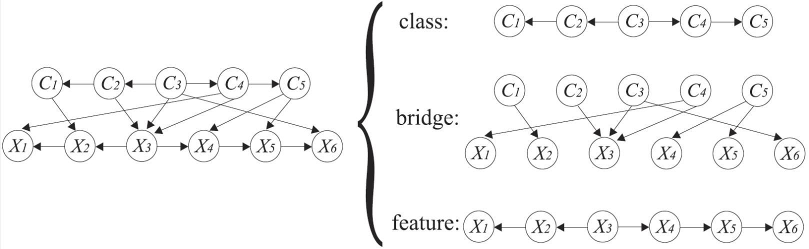
$$h : (PDQ_1, ..., PDQ_{39}) \rightarrow (EQ_1, ..., EQ_5)$$

Borchani, Bielza, Martínez-Martín, Larrañaga (2012). Multidimensional Bayesian network classifiers applied to predict the European quality of life-5 dimensions (EQ-5D) from the 39-item Parkinson's disease questionnaire (PDQ-39), ***Journal of Biomedical Informatics***, accepted

Multi-dimensional classification for EQ-5D health states from PDQ-39 in Parkinson

Multi-dimensional Bayesian network classifier (MBC)

- The set of variables $\mathcal{V}$ is partitioned into:
  - $\mathcal{V}_{\mathcal{C}} = \{C_1, ..., C_d\}$ of class variables and
  - $\mathcal{V}_{\mathcal{X}} = \{X_1, ..., X_m\}$ of feature variables



Most probable explanation (MPE)

$$(c_1^*, ..., c_d^*) = \max_{c_1,...,c_d} p(C_1 = c_1, ..., C_d = c_d | X_1 = x_1, ..., X_m = x_m)$$

Bielza, Li, Larrañaga (2011). Multi-dimensional classification with Bayesian networks, ***International Journal of Approximate Reasoning***, 52(6), 705-727

Multi-dimensional classification for EQ-5D health states from PDQ-39 in Parkinson

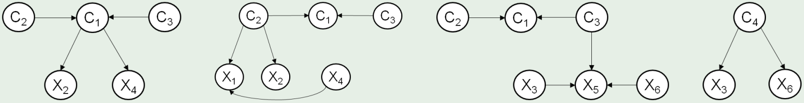
Four MBC learning algorithms

1. Markov blanket - Multi-dimensional Bayesian classifier (MB-MBC) [Borchani et al., 2011]

Multi-dimensional classification for EQ-5D health states from PDQ-39 in Parkinson

Four MBC learning algorithms

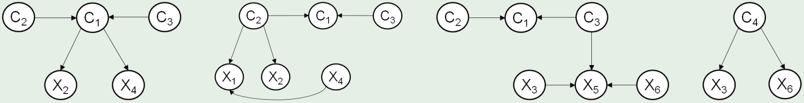
1. Markov blanket – Multi-dimensional Bayesian classifier (MB–MBC) [Borchani et al., 2011]

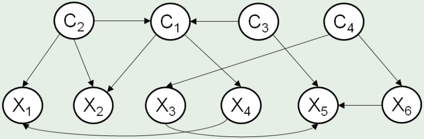


Multi-dimensional classification for EQ-5D health states from PDQ-39 in Parkinson

Four MBC learning algorithms

1. Markov blanket – Multi-dimensional Bayesian classifier (MB–MBC) [Borchani et al., 2011]





Multi-dimensional classification for EQ-5D health states from PDQ-39 in Parkinson

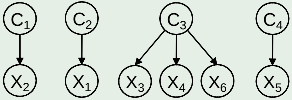
Four MBC learning algorithms

1. Markov blanket - Multi-dimensional Bayesian classifier (MB-MBC) [Borchani et al., 2011]
2. Class-Bridge - Multi-dimensional Bayesian classifier (CB-MBC) [Borchani et al., 2010]

Multi-dimensional classification for EQ-5D health states from PDQ-39 in Parkinson

Four MBC learning algorithms

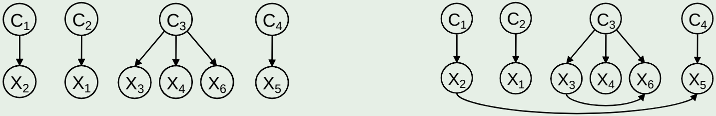
1. Markov blanket - Multi-dimensional Bayesian classifier (MB-MBC) [Borchani et al., 2011]
2. Class-Bridge - Multi-dimensional Bayesian classifier (CB-MBC) [Borchani et al., 2010]



Multi-dimensional classification for EQ-5D health states from PDQ-39 in Parkinson

Four MBC learning algorithms

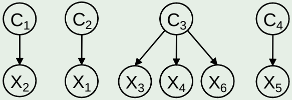
1. Markov blanket - Multi-dimensional Bayesian classifier (MB-MBC) [Borchani et al., 2011]
2. Class-Bridge - Multi-dimensional Bayesian classifier (CB-MBC) [Borchani et al., 2010]

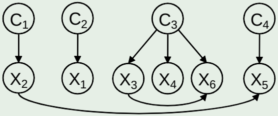


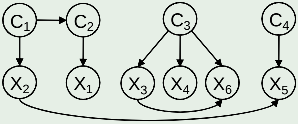
Multi-dimensional classification for EQ-5D health states from PDQ-39 in Parkinson

Four MBC learning algorithms

1. Markov blanket − Multi-dimensional Bayesian classifier (MB−MBC) [Borchani et al., 2011]
2. Class-Bridge − Multi-dimensional Bayesian classifier (CB−MBC) [Borchani et al., 2010]



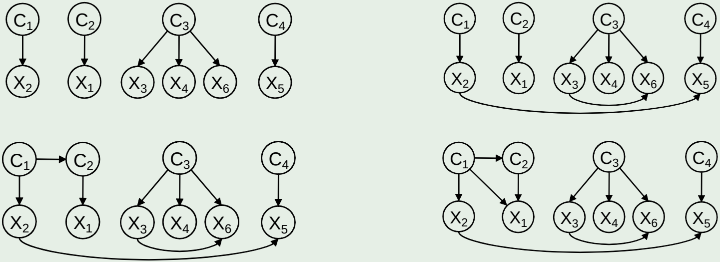




Multi-dimensional classification for EQ-5D health states from PDQ-39 in Parkinson

Four MBC learning algorithms

1. Markov blanket – Multi-dimensional Bayesian classifier (MB–MBC) [Borchani et al., 2011]
2. Class-Bridge – Multi-dimensional Bayesian classifier (CB–MBC) [Borchani et al., 2010]



Multi-dimensional classification for EQ-5D health states from PDQ-39 in Parkinson

Four MBC learning algorithms

1. Markov blanket - Multi-dimensional Bayesian classifier (MB−MBC) [Borchani et al., 2011]
2. Class−Bridge - Multi−dimensional Bayesian classifier (CB−MBC) [Borchani et al., 2010]
3. Independent Markov blanket classifiers with HITON algorithm (Indep−MB−HITON) [Aliferis et al., 2010]

Multi-dimensional classification for EQ-5D health states from PDQ-39 in Parkinson

Four MBC learning algorithms

1. Markov blanket – Multi-dimensional Bayesian classifier (MB–MBC) [Borchani et al., 2011]
2. Class-Bridge – Multi-dimensional Bayesian classifier (CB–MBC) [Borchani et al., 2010]
3. Independent Markov blanket classifiers with HITON algorithm (Indep–MB–HITON) [Aliferis et al., 2010]
4. Independent Markov blanket classifiers with PC algorithm (Indep–MB–PC) [Le and Doctor, 2011]

Multi-dimensional classification for EQ-5D health states from PDQ-39 in Parkinson

Four MBC learning algorithms

1. Markov blanket - Multi-dimensional Bayesian classifier (MB-MBC) [Borchani et al., 2011]
2. Class-Bridge - Multi-dimensional Bayesian classifier (CB-MBC) [Borchani et al., 2010]
3. Independent Markov blanket classifiers with HITON algorithm (Indep-MB-HITON) [Aliferis et al., 2010]
4. Independent Markov blanket classifiers with PC algorithm (Indep-MB-PC) [Le and Doctor, 2011]

- Borchani, Bielza, Larrañaga (2011). *Probabilistic Problem Solving in Biomedicine. Workshop in the 13th Conference on Artificial Intelligence in Medicine*, 29-40
- Borchani, Bielza, Larrañaga (2010). *Proc. of the 5th Workshop on Probabilistic Graphical Models*, 25-32
- Aliferis et al. (2010). *Journal of Machine Learning Research*, 11, 235-284
- Le and Doctor (2011). *Medical Care*, 49(5), 451-460

Multi-dimensional classification for EQ-5D health states from PDQ-39 in Parkinson

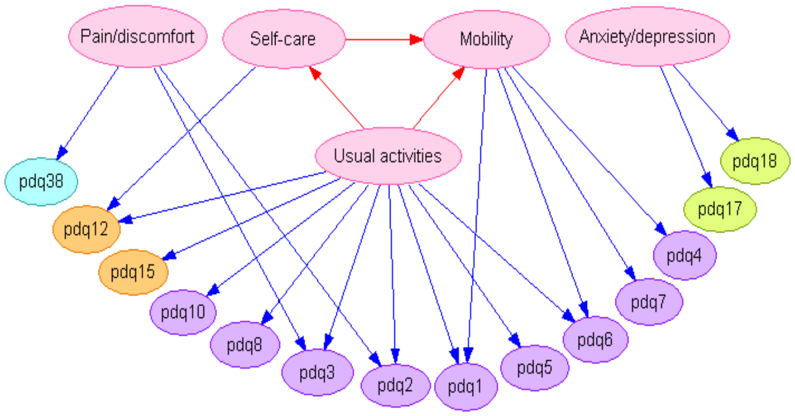
Parkinson 488 patients. Estimated accuracies over 5-fold cross-validation

Method	Mean accuracy	Global accuracy
`MB-MBC`	**0.7119** ± **0.0338**	0.2030 ± 0.0718
`CB-MBC`	0.6807 ± 0.0285	0.1865 ± 0.0429
`Indep-MB-HITON`	0.7009 ± 0.0427	**0.2051** ± **0.0835**
`Indep-MB-PC`	0.6587 ± 0.0636	0.1867 ± 0.0937
`MNL`	0.6926 ± 0.0430	0.1802 ± 0.0713
`OLS`	0.4201 ± 0.0252	0.0123 ± 0.0046
`CLAD`	0.4254 ± 0.0488	0.0143 ± 0.0171

- Mean accuracy over the d class variables: $Acc_m = \frac{1}{d}\sum_{i=1}^{d}\frac{1}{N}\sum_{l=1}^{N}\delta(\hat{c}_{li}, c_{li})$
- Global accuracy over the d-dimensional class variable: $Acc_g = \frac{1}{N}\sum_{l=1}^{N}\delta(\hat{\mathbf{c}}_l, \mathbf{c}_l)$

Multi-dimensional classification for EQ-5D health states from PDQ-39 in Parkinson

MB–MBC graphical structure



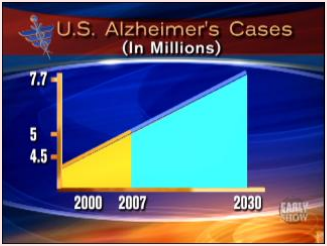
Outline

Knowledge discovery in Alzheimer's disease

Alzheimer's disease

- Primarily affects the elderly and manifests through memory disorders, cognitive decline and loss of autonomy

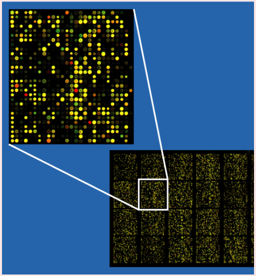
Alois Alzheimer (1864-1915)

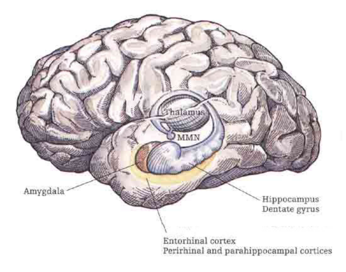


- In 2011, 33.9 million cases worldwide. Predicted to affect 1 in 85 people by 2050
- Every 70 seconds, someone is diagnosed with Alzheimer's
- Seventh-leading cause of death

Knowledge discovery in Alzheimer's disease

Alzheimer's disease and DNA microarrays





- Idea in [Small et al., 2005]: microarray data selectively from the brain site most
  - vulnerable to AD to maximize expression differences between AD and controls: entorhinal cortex (EC)
- 6 AD brains + 6 control brains ⇒ 12 tissue samples and 7,610 variables

Small et al. (2005). Model-guided microarray implicates the retromer complex in Alzheimer's disease, *Annals of Neurology*, 58(6), 909-919

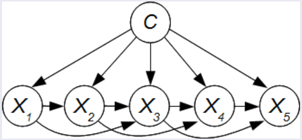
Knowledge discovery in Alzheimer's disease

⇒ Re-analyze the data differently to gain robustness (small sample size!)

⇒ Find out explicit new (or validate old) biological relationships and genes not previously reported

Reliable-kDB classifier with robust gene interactions

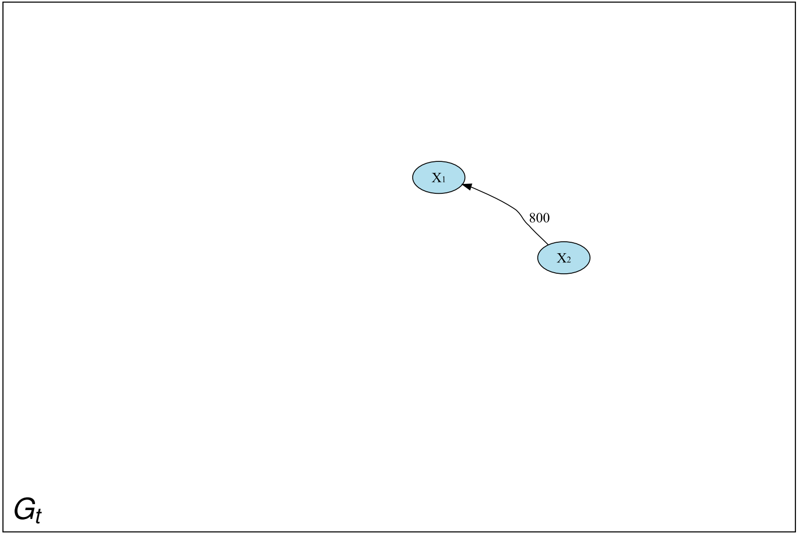
- Learn a Bayesian network classifier. We use kDB structures with at most k parents (excluding the class)



- Induce many kDB by a resampling method (bootstrap) with an inner FSS
- Output a network with those arcs above a reliability threshold t: arcs occurring $\geq t$ times are retained
- Approach is a consensus feature selection on the final gene interaction network

Armañanzas, Larrañaga, Bielza (2012). Ensemble transcript interaction networks: A case study on Alzheimer's disease, ***Computer Methods and Programs in Biomedicine***, 108, 442-450

Reliable-kDB classifier –An example



t

800 <

798

790

⋮

700

600

500

400

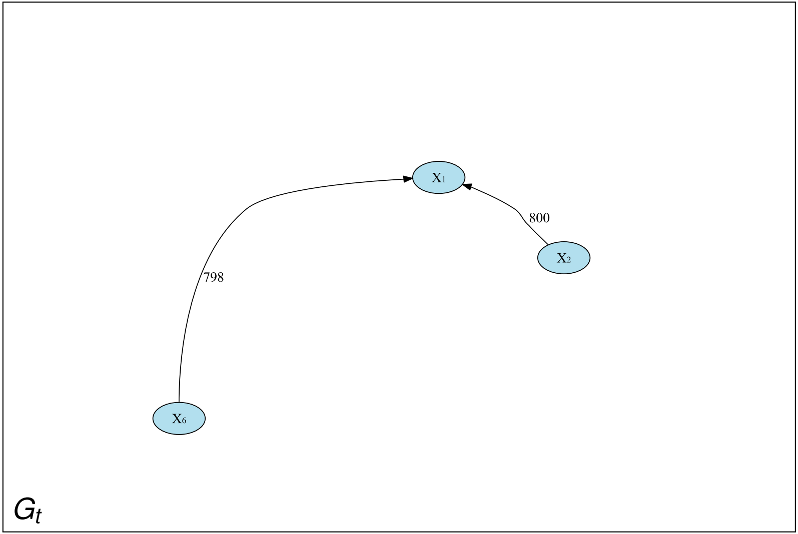
⋮

100

⋮

1

Reliable-kDB classifier –An example



t

800

798 $<$

790

⋮

700

600

500

400

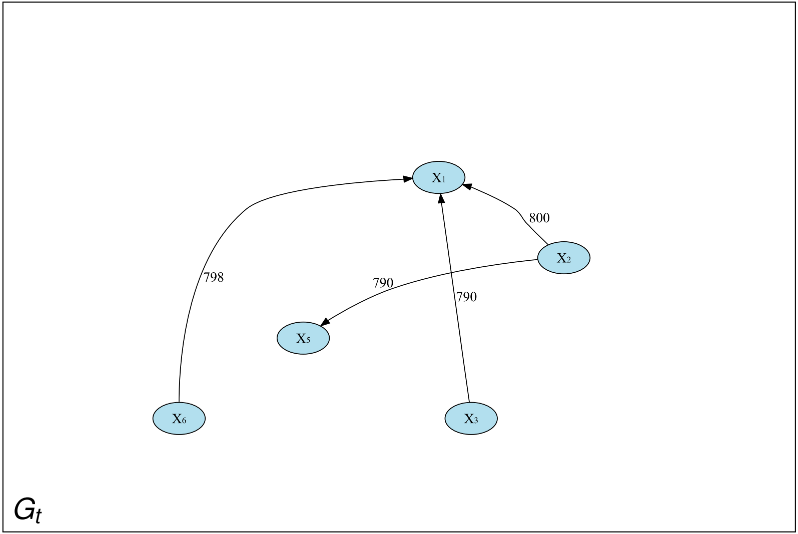
⋮

100

⋮

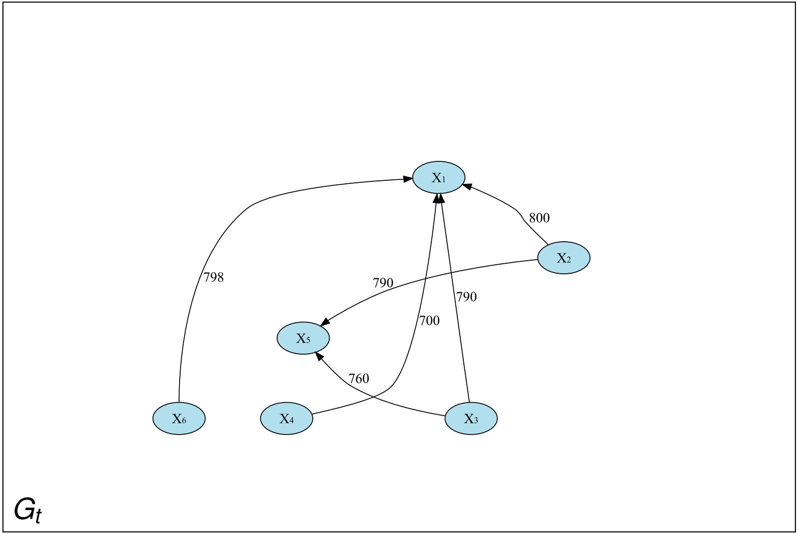
1

Reliable-kDB classifier –An example



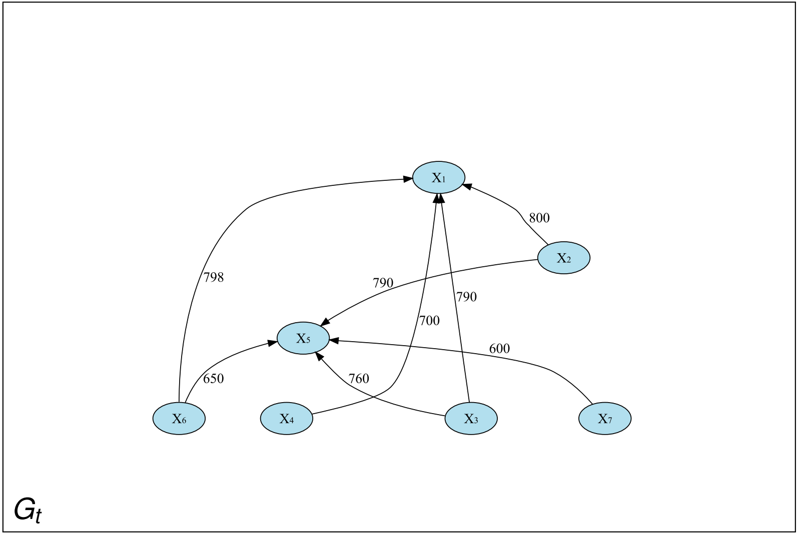
t
800
798
790 $<$
⋮
700
600
500
400
⋮
100
⋮
1

Reliable-kDB classifier –An example



t
800
798
790
⋮
700 $<$
600
500
400
⋮
100
⋮
1

Reliable-kDB classifier –An example



t

800

798

790

⋮

700

600 $<$

500

400

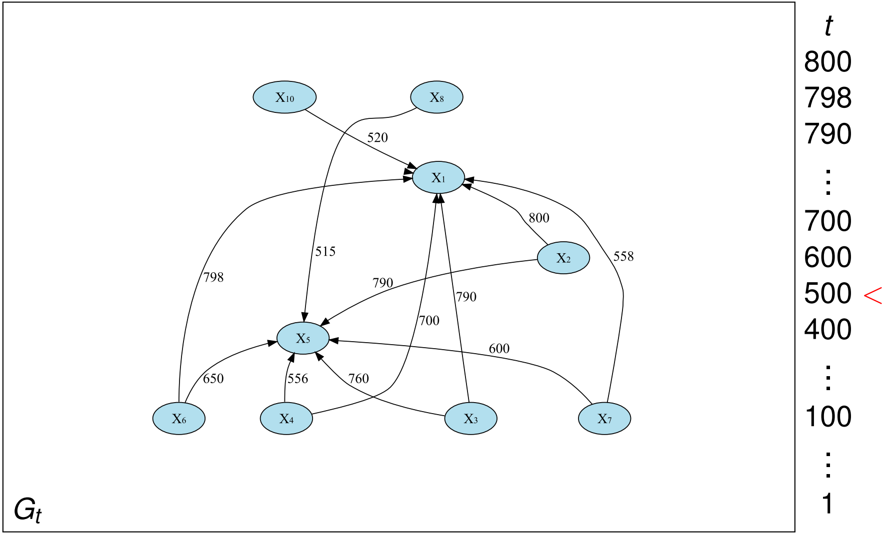
⋮

100

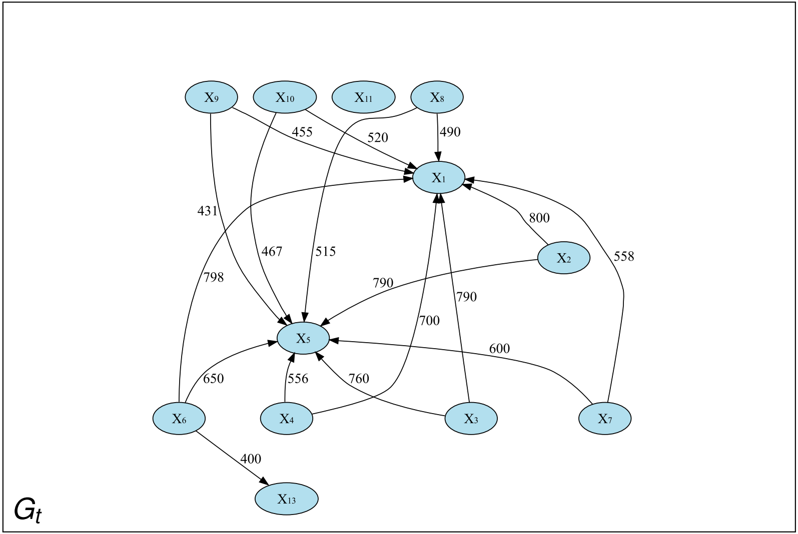
⋮

1

Reliable-kDB classifier –An example



Reliable-kDB classifier –An example



t

800

798

790

⋮

700

600

500

400 <

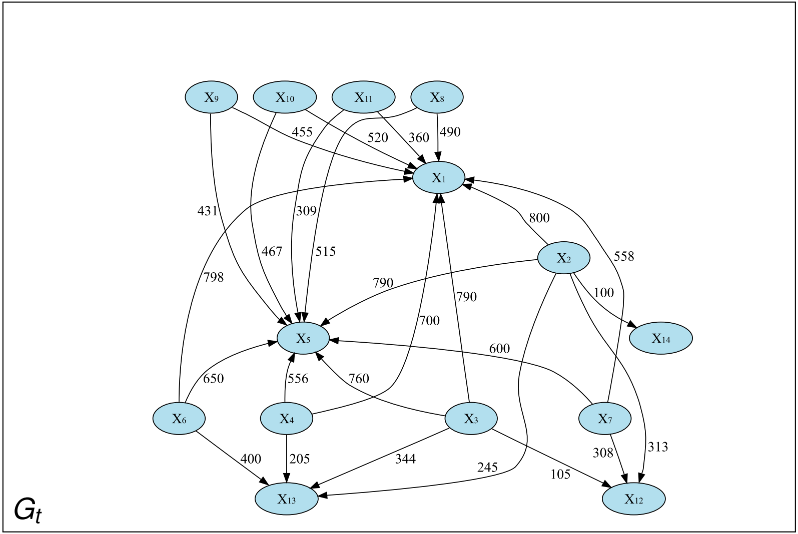
⋮

100

⋮

1

Reliable-kDB classifier –An example



t

800

798

790

⋮

700

600

500

400

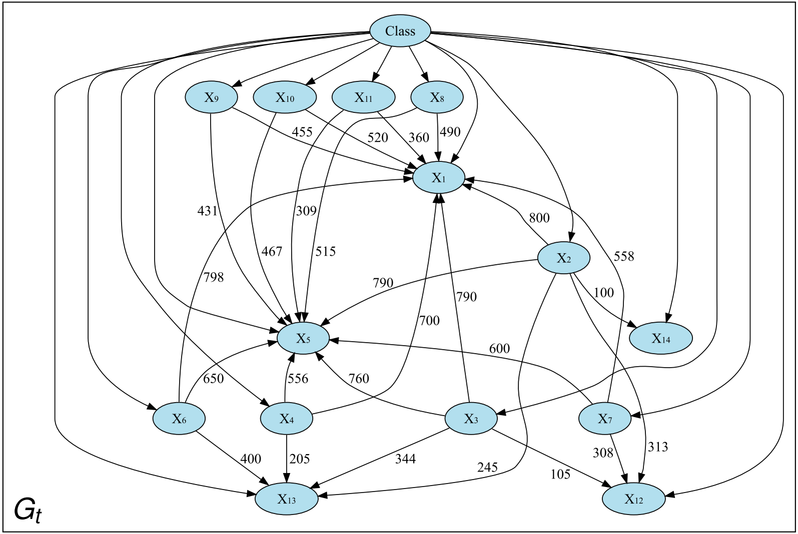
⋮

100 <

⋮

1

Reliable-kDB classifier –An example



t

800

798

790

⋮

700

600

500

400

⋮

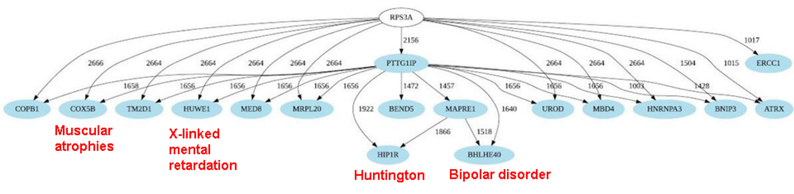
100 <

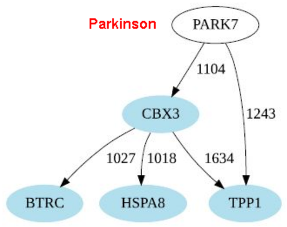
⋮

1

Knowledge discovery in Alzheimer’s disease

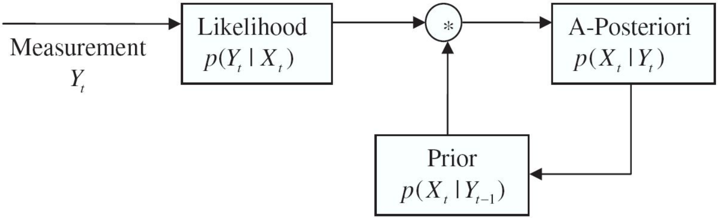
EC-AD vs. EC-Control (12 samples) with $k = 4$, $B = 10000$, $t = 1000$





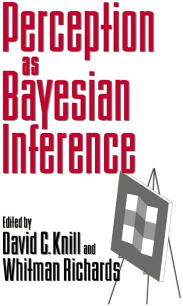
Outline

The Bayesian brain

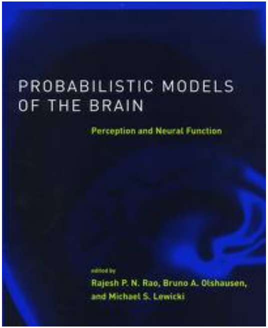


- Kording (2007). Decision theory: what “should” the nervous system do? *Science*, 318, 606-610
- Colombo, Seriès (2012). Bayes in the brain. On Bayesian modelling in neuroscience, *The British Journal for the Philosophy of Sciences*, in press

The Bayesian brain: books



Cambridge University Press (1996)



The Mit Press (2002)

The Mit Press (2007)

Alan Turing and Bayesian statistics

Biometrika (1979), **66**, 2, *pp*, 393–6
Printed in Great Britain

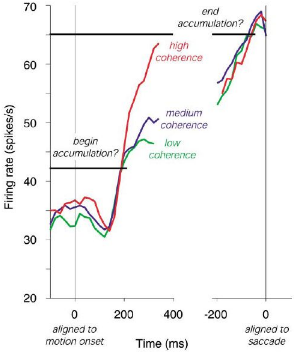
Studies in the History of Probability and Statistics. XXXVII
A. M. Turing's statistical work in World War II

BY I. J. GOOD

- Weight of evidence, or log (Bayes) factor, in favour of a hypothesis, H, provided by evidence, E: $\log \frac{P(E|H)}{P(E|\overline{H})}$
- Banburismus algorithm:
  - Weight of evidence (H possible configurations of the Enigma machine, and E matches under the hypothesis)
  - Update the weight of evidence with more evidence
  - A decision rule (for deciding between two hypotheses, H_1 and H_2)

Alan Turing and Neuroscience

Enigma machine



Banburismus in the brain

- Gold, Shadlen (2002). Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward, *Neuron*, 36, 299-308
- Larrañaga, Bielza, DeFelipe (2012). Alan Turing and neuroscience, ***Investigación y Ciencia***, in press

Outline

Bayesian networks in neuroscience

Challenging machine learning problems in modeling the brain

- JOINT PROBABILITY DISTRIBUTION: Bayesian networks (dendritic morphology)
- SEMI-SUPERVISED WITH CLASS DISCOVERY: EM based subspace clustering (new types of neurons)
- SUPERVISED CLASSIFICATION WITH PROBABILISTIC LABELS: Bayesian classifiers (neuron class with probabilistic labels)
- CONSENSUS OF PROBABILISTIC MODELS: Bayesian networks (a neuroscientist $\equiv$ a model)
- BAYESIAN CLASSIFIERS: Selective naive Bayes (dementia development in Parkinson's disease)
- MULTI-DIMENSIONAL CLASSIFICATION: multi-dimensional Bayesian classifiers (from PDQ-39 to EQ-5D in Parkinson's disease)
- BOOTSTRAP FOR RELIABLE MODELS: k-DB Bayesian classifiers (knowledge discovery in Alzheimer's disease)

Thanks to

- Cajal Blue Brain project

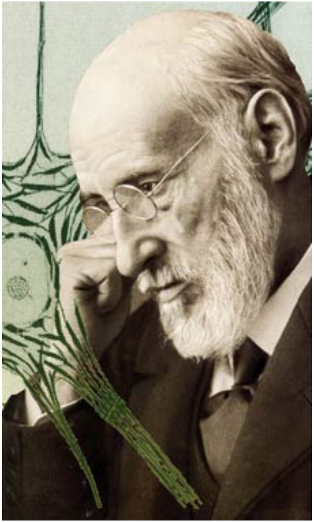
Technical University of Madrid

The Cajal Institute

- Consolider Ingenio 2010-CSD2007-00018 project
- TIN2010-20900-C04-04 project

Judea Pearl and Santiago Ramón y Cajal

BAYESIAN NETWORKS IN NEUROSCIENCE



Bayesian Networks in Neuroscience

Concha Bielza, Pedro Larrañaga

Computational Intelligence Group
Artificial Intelligence Department
Technical University of Madrid, Spain

PGM'12
Granada, September 19, 2012