Updating Strategies and Speed/Accuracy trade-offs in Variable Selection Algorithms for Multivariate Data Analysis

A. PEDRO DUARTE SILVA

FEG / CEGE
UNIVERSIDADE CATÓLICA PORTUGUESA
CENTRO REGIONAL DO PORTO

Supported by: FEDER / POCI 2010

Ciência.Inovação 2010 Programa Operacional Ciência e Inovação 2010
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA E DA TECNOLOGIA

Speed and Accuracy in Variable Selection

OUTLINE

1. CRITERIA FOR VARIABLE SELECTION

1.1. LINEAR MODELS WITH MULTIPLE RESPONSES

1.2. A LINEAR HYPOTHESIS FRAMEWORK

2. ALGORITHMS BASED ON SSCP MATRICES

2.1. UPDATING STRATEGIES

2.2 COMPUTATIONAL EFFORT

2.3. ERROR BOUNDS AND STABILITY CONDITIONS

3. AN ALGORITHM BASED ON THE ORIGINAL DATA

4. FINAL REMARKS

Speed and Accuracy in Variable Selection

LINEAR MODELS WITH MULTIPLE RESPONSES

$$\mathbf{Y = X\,B + U} \qquad \mathbf{T = S_{XX}} \qquad \mathbf{E = S_{XY}\,S_{YY}^{-1}\,S_{YX}} \qquad \mathbf{H = T - E}$$

$\mathrm{ccr}_i^2 = \mathrm{Eigval}_i(T^{-1}H)$ **(i = 1, ..., r)** **r = rank(H) =** ***min*** **(ncol(X),ncol(Y))**

Comparison Criteria:

Multivariate Indices

$$\mathbf{ccr}_1^2 = \frac{\lambda_1}{1+\lambda_1}$$

$\left(\textit{max}\ \mathrm{ccr}_1^2 \Leftrightarrow \textit{max}\ \lambda_1 = \mathrm{Eigval}_1(T^{-1}E) \right)$

→ **Roy criterion**

$$\tau^2 = 1 - \left(\prod_{i=1}^{r} (1 - \mathrm{ccr}_i^2) \right)^{1/r} = 1 - \Lambda^{1/r}$$

$$\left(\max\ \tau^2 = \min\ \Lambda = \frac{\det(W)}{\det(T)} \right)$$

→ **Wilks criterion**

$$\varsigma^2 = 1 - \frac{r}{\sum_{i=1}^{r} (1-\mathrm{ccr}_i^2)^{-1}} = \frac{v}{r+v}$$

$\left(\textit{max}\ \varsigma^2 \Leftrightarrow \textit{max}\ v = \textit{tr}\ E^{-1}H \right)$

→ **Lawley-Hotelling criterion**

$$\xi^2 = \frac{\sum_{i=1}^{r} \mathrm{ccr}_i^2}{r} = \frac{u}{r}$$

$\left(\textit{max}\ \xi^2 \Leftrightarrow \textit{max}\ u = \textit{tr}\ T^{-1}H \right)$

→ **Bartlett-Pillai criterion**

Speed and Accuracy in Variable Selection

A LINEAR HYPOTHESIS FRAMEWORK

$X = A\,\Psi + U$ H0: $C\,\Psi = 0$

$\rightarrow$ **SELECT COLUMNS OF X IN ORDER TO EXPLAIN H1**

PARTICULAR CASES:

(i) LINEAR DISCRIMINANT ANALYSIS

$$A = [1_g] \quad \Psi = [\mu_g] \qquad C = \begin{bmatrix} 1 & -1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 1 & 0 & \dots & -1 \end{bmatrix}$$

(ii) MULTI-WAY MANOVA/MANCOVA EFFECTS

$$\Omega = \mathcal{R}(A) \qquad \omega = \mathcal{R}(A) \cap \mathcal{N}(C) \qquad r = \dim(\Omega) - \dim(\omega)$$

$$ccr_i^2 = \text{Eigval}_i(T^{-1}H) \qquad T = X'(I - P_\omega)X \qquad H = X'(P_\Omega - P_\omega)X$$

Speed and Accuracy in Variable Selection

ALGORITHMS BASED ON SSCP MATRICES

FURNIVAL'S ALGORITHM FOR LINEAR REGRESSION (1971):

$$y = X'\hat{\beta} + e \qquad \textit{min}\ s_e^2 = s_y^2 - S_{yX} S_{XX}^{-1} S_{Xy}$$

$$A = \begin{bmatrix} S_{XX} & S_{Xy} \\ S_{Xy} & s_y^2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & S_{1y} \\ S_{21} & S_{22} & S_{2y} \\ S_{y1} & S_{y2} & s_y^2 \end{bmatrix}$$

GAUSSIAN ELIMINATION PIVOTS

$$X = [\, X_1 \mid X_2 \,]$$

$$\begin{bmatrix} \dots & -S_{11}^{-1}S_{12} & -S_{11}^{-1}S_{1y} \\ -S_{21}S_{11}^{-1} & S_{22} - S_{21}S_{11}^{-1}S_{12} & S_{2y} - S_{21}S_{11}^{-1}S_{1y} \\ -S_{y1}S_{11}^{-1} & S_{y2} - S_{y1}S_{11}^{-1}S_{12} & \mathbf{s_y^2 - S_{y1}S_{11}^{-1}S_{1y}} \end{bmatrix}$$

Speed and Accuracy in Variable Selection

ALGORITHMS BASED ON SSCP MATRICES

COMMENTS

(i) ONLY THE RIGHT-LOWER CORNER OF **A** NEEDS TO BE UPDATED AT EACH STEP

1/2 OF THE PIVOTS JUST HAVE TO COMPUTE s^2_e

1/4 OF THE PIVOTS UPDATE (2*2) SYMMETRIC SUBMATRICES

…

ONE SINGLE PIVOT NEEDS TO UPDATE A (P*P) SYMMETRIC MATRIX

$\Rightarrow$ TOTAL NUMBER OF FLOATING POINT OPERATIONS:

$6(2^p) - (1/2)p^2 - (7/2)p - 6$ multiplications/divisions

$\underline{4(2^p) - (1/2)p^2 - (5/2)p - 4}$ additions/subtractions

$10(2^p) - p^2 - 6p - 10$ flops

Speed and Accuracy in Variable Selection

ALGORITHMS BASED ON SSCP MATRICES

COMMENTS

(ii) THE SAME PROCEDURE CAN BE APPLIED MOVING BACKWARDS

FROM $\begin{bmatrix} -S_{XX}^{-1} & -S_{XX}^{-1} S_{Xy} \\ -S_{yX} S_{XX}^{-1} & s_e^2 \end{bmatrix}$ TO $\Longrightarrow$ $\begin{bmatrix} \ldots & \ldots \\ \ldots & s_y^2 \end{bmatrix}$

$\Rightarrow$ BRANCH AND BOUND ALGORITHMS (FURNIVAL AND WILSON 1974)

WORST CASE (NO PRUNING) TIME COMPLEXITY

$6(2^p) + O(p^2)$	multiplications/divisions
$\underline{4(2^p) + O(p^2)}$	additions/subtractions
$10(2^p) + O(p^2)$	flops

Speed and Accuracy in Variable Selection

ALGORITHMS BASED ON SSCP MATRICES

COMMENTS

(iii) AND REMAINS VALID ON APPLIED TO:

- QUADRATIC FORMS OF THE TYPE $v_1' \Phi_{11}^{-1} v_1$
- DETERMINANTS OF THE TYPE $\det(\Phi_{11})$
- SUMS AND RATIOS OF THE ABOVE CRITERIA

$$\Phi = \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix}$$

SYMMETRIC, NON-SINGULAR AND POSITIVE-DEFINITE

WORST CASE TIME COMPLEXITY:

→ Lawley-Hotelling criterion

$$v = tr\left(E^{-1}H\right) = tr\left[E^{-1}\left(\sum_{i=1}^{r} h_i h_i'\right)\right] = \sum_{i=1}^{r} h_i' E^{-1} h_i$$

$(5+5r)(2^p) + O(p^2)$ flops

→ Bartlett-Pillai criterion

$$u = tr\left(T^{-1}H\right) = tr\left[T^{-1}\left(\sum_{i=1}^{r} h_i h_i'\right)\right] = \sum_{i=1}^{r} h_i' T^{-1} h_i$$

$(5+5r)(2^p) + O(p^2)$ flops

→ Wilks criterion

$$\Lambda = \det(E) / \det(T)$$

$12(2^p) + O(p^2)$ flops

Speed and Accuracy in Variable Selection

ALGORITHMS BASED ON SSCP MATRICES

BASIC ALGORITHM $\Leftrightarrow$ L D L' DECOMPOSITIONS OF SYMMETRIC POSITIVE-DEFENITE MATRICES, M_{11}, BY GAUSSIAN ELIMINATION

IF # X_1 = K $\rightarrow$

$$\nu_1' \Phi_{11}^{-1} \nu_1 = \alpha - D_{11}(k+1,K+1)$$

$$\det(\Phi_{11}) = \det(M_{11}) = \prod_{i=1}^{k} D_{ii}$$

ERROR ANALYSIS

COMPUTED $\hat{L}\,\hat{D}\,\hat{L}' = M_{11} + \Delta\, M_{11}$

$$\|\Delta M_{11}\|_2 \leq \gamma_k \|\,|\hat{L}|\,\hat{D}|\hat{L}'|\,\|_2$$

$$\gamma_l \leq \frac{l\,u}{1-l\,u}$$

l = dim(M_{11})

u: = unit roundoff

AND

$$\|D-\hat{D}\|_F \leq \frac{\|M_{11}\|_2\,\|M_{11}^{-1}\|_2\,\|\Delta M_{11}\|_F}{1-\|M_{11}^{-1}\|_2\,\|\Delta M_{11}\|_2}$$

(BARRLUND 1991)

$$\Rightarrow\ |D(j)-\hat{D}(j)| \leq \frac{k^{3/2}\gamma_k(1-k\gamma_k)^{-1}\kappa_2(M_{11})\|M_{11}\|_2}{1-k(1-k\gamma_k)^{-1}\kappa_2(M_{11})}$$

$$\kappa_2(M_{11}) = \|M_{11}\|_2\,\|M_{11}^{-1}\|_2$$

Speed and Accuracy in Variable Selection

ALGORITHMS BASED ON SSCP MATRICES

ERROR ANALYSIS

Lawley-Hotelling criterion

$$|v - \hat{v}| \le \frac{(p+1)^{3/2}\gamma_{p+1}}{1-(p+1)\gamma_{p+1}} \sum_{i=1}^{r} \frac{\kappa_2(M_i)\|M_i\|_2}{1-(p+1)\left(1-(p+1)\gamma_{p+1}\right)^{-1}\kappa_2(M_i)}$$

$$M_i = \begin{bmatrix} E & h_i \\ h_i' & \alpha_i \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} -E^{-1} & -E^{-1}h_i \\ -h_i'E^{-1} & \alpha_i - h_i'E^{-1}h_i \end{bmatrix}$$

Bartlett-Pillai criterion

$$|u - \hat{u}| \le \frac{(p+1)^{3/2}\gamma_{p+1}}{1-(p+1)\gamma_{p+1}} \sum_{i=1}^{r} \frac{\kappa_2(N_i)\|N_i\|_2}{1-(p+1)\left(1-(p+1)\gamma_{p+1}\right)^{-1}\kappa_2(N_i)}$$

$$N_i = \begin{bmatrix} T & h_i \\ h_i' & \beta_i \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} -T^{-1} & -T^{-1}h_i \\ -h_i'E^{-1} & \beta_i - h_i'T^{-1}h_i \end{bmatrix}$$

Speed and Accuracy in Variable Selection

ALGORITHMS BASED ON SSCP MATRICES

ERROR ANALYSIS

Wilks criterion

$$|\Lambda - \hat{\Lambda}| \leq \frac{p^{3/2}\gamma_p}{1 - p\gamma_p} \frac{\kappa_2(M)\kappa_2(N)\,\|M\|_2\,\|N\|_2}{1 - p\left(1 - p\gamma_p\right)^{-1}\left(\kappa_2(M) + \kappa_2(N)\right) + \kappa_2(M)\kappa_2(N)} \; *$$

$$* \sum_{i=1}^{r}\left(\frac{1}{Egval_i(M)} + Egval_i(N)\right) + O(u^2)$$

M = E or $-E^{-1}$ **N = T or $-T^{-1}$**

Speed and Accuracy in Variable Selection

AN ALGORITHM BASED ON THE ORIGINAL DATA

$X = A\,\Psi + U \qquad H0:\ C\,\Psi = 0$

$\Omega = \mathcal{R}(A) \qquad \omega = \mathcal{R}(A) \cap \mathcal{N}(C) \qquad r = \dim(\Omega) - \dim(\omega)$

$ccr_i^2 = \mathrm{Eigval}_i\,(T^{-1}H) \qquad T = X'(I - P_\omega)\,X \qquad H = X'(P_\Omega - P_\omega)\,X$

<u>KEY INSIGHT:</u>

THE COMPUTATION OF ccr_i^2 CAN ALWAYS BE BASED ON A GENERALIZED SINGULAR VALUE PROBLEM

WRITING: $H = X_H' X_H \qquad E = X_E' X_E$

$X_H = U_H R_H Q' \qquad X_E' = U_E R_E Q' \qquad R_H = D_{H/E} R_E$

THEN $ccr_i^2 = (D_{H/E}(i,i))^2 \,/\, [1 + (D_{H/E}(i,i))^2]$

Speed and Accuracy in Variable Selection

AN ALGORITHM BASED ON THE ORIGINAL DATA

<u>INITIALIZATION STEP</u>

FIND X_H AND X_E DIRECTLY FROM X, A AND C

FROM THE SVD's

$$A = U \Sigma V' = [U_s \; \tilde{U}] \, \Sigma \, [V_s \; \tilde{V}]'$$

$$C V_S \Sigma_S^2 V'_S C' = F D F'$$

WE GET

$$X_E = \tilde{U}' X$$

$$X_H = D^{-1/2} F' C V_S \Sigma_S^{-1} U_S' X$$

Speed and Accuracy in Variable Selection

AN ALGORITHM BASED ON THE ORIGINAL DATA

<u>ITERATION STEP</u>

$X_H = U_H R_H Q'$

$X_E' = U_E R_E Q'$

$R_H = D_{H/E} R_E$

$$[X_{H1} \mid X_H(.,k+1)] = U\,[R_H \mid U'_H X_H(.,k+1)] \begin{bmatrix} Q' & 0 \\ 0 & 1 \end{bmatrix}$$

$$[X_{E1} \mid X_E(.,k+1)] = U\,[R_E \mid U'_E X_E(.,k+1)] \begin{bmatrix} Q' & 0 \\ 0 & 1 \end{bmatrix}$$

BUT

$[R_H \mid U'_H X_H(.,k+1)]$ MIGHT NOT BE TRIANGULAR

AND

$[R_E \mid U'_E X_E(.,k+1)]$ MIGHT NOT BE PARALLEL

Speed and Accuracy in Variable Selection

AN ALGORITHM BASED ON THE ORIGINAL DATA

<u>ITERATION STEP</u>

RESTORING TRIANGULAR STRUCTURE

- **HOUSEHOLDER TRANSFORMATIONS ON**

$$\left[R_H \mid U'_H X_H(.,k+1)\right] \quad \text{AND} \quad \left[R_E \mid U'_E X_E(.,k+1)\right]$$

RESTORING PARALLELISM

- **PAIGE'S (1986) ELEMENTARY 2*2 ROTATIONS**

TIME COMPLEXITY OF EACH ITERATION: $O(s*k + r*k + k^2)$

TOTAL TIME COMPLEXITY: $O(p^2 2^P)$

Speed and Accuracy in Variable Selection

FINAL REMARKS

NUMERICAL ACCURACY

IN MOST PRATICAL PROBLEMS THE DATA DOES NOT EXHIBIT A PATTERN OF MULTICORRELATION STRONG ENHOUGH TO MAKE THE STABILITY OF ANY OF THE ALGORITHMS A RELEVANT CONCERN

WHEN SEVERE MULTICORRELATION IS PRESENT NONE OF THE ALGORITHMS CAN GIVE RELIABLE RESULTS

OFTEN STATISTICAL REASONS RECOMMEND THAT MULTICOLINEAR SUBSETS SHOULD BE AVOID, LONG BEFORE NUMERICAL ACCURACY BECAMES AN ISSUE

Speed and Accuracy in Variable Selection

FINAL REMARKS

<u>COMPUTATIONAL EFFORT</u>

WHEN THE NUMBER OF ORIGINAL VARIABLES IS MODERATE ALL ALGORITHMS ARE FAST ENOUGH SO THAT COMPUTATIONAL EFFORT IS NOT A RELEVANT CONCERN

WHEN THE NUMBER OF ORIGINAL VARIABLES IS LARGE NONE OF THE ALGORITHMS CAN PROVE OPTIMAL RESULTS IN A REASONABLE TIME

WHEN TRUE OPTIMALITY CANNOT BE PROVEN THERE ARE NEVERTHELESS MANY HEURISTIC METHODS THAT ARE ABLE TO QUICKLY IDENTIFY GOOD (OFTEN THE BEST) VARIABLE SUBSETS EVEN IN PROBLEMS WITH A VERY LARGE NUMBER OF ORIGINAL VARIABLES

References

Barrlund, A. (1991). Pertubation Bounds for the LDL^H and LU Decompositions. *BIT* **31**: 358-363.

Duarte Silva, A.P. (2001). Efficient Variable Screening for Multivariate Analysis. *Journal of Multivariate Analysis* **76**, 35-62.

Furnival, G.M. (1971). All Possible Regressions with Less Computation. *Technometrics* **13**: 403-408.

Furnival, G.M. & Wilson, R.W. (1974). Regressions by Leaps and Bounds. *Technometrics* **16**: 499-511.

Higham, N.J. (2002). *Accuracy and Stability of Numerical Algorithms. 2nd Ed.* SIAM Philadelphia, PA.

Paige, C.C. & Saunders, M.A. (1981). Towards a Generalized Singular Value Decomposition. *SIAM Journal of Numerical Analysis* **18**: 398-405.

Paige, C.C. (1986) Computing the Generalized Singular Value Decomposition. *SIAM Journal of Scientific and Statistical Computing* **7**: 1126-1146.