Visualizing Correlations in β

Dr. Scott, Stat 410

October 13, 2005

For our model $\qquad Y = X\beta + \epsilon$

$$\min_{\beta} SS(\beta) = \epsilon^t \epsilon = (Y - X\beta)^T (Y - X\beta)$$

$$= Y^t Y - 2\beta^t X^t Y + \beta^t X^t X \beta$$

The least-squares coefficient solves

$$\underline{0} = \nabla_\beta SS = \underline{0} - 2X^t Y + 2X^t X \beta$$

$$\Rightarrow \quad \hat{\beta} = (X^t X)^{-1} X^t Y$$

$$\text{Hessian} = \nabla \nabla^t SS = 2X^t X$$

which is pos. def. $\Rightarrow$ minimizer!!

Statistical Accuracy View

$$
\begin{aligned}
E\widehat{\beta} &= (X^t X)^{-1} X^t \, E(Y) \\
&= (X^t X)^{-1} X^t [X\beta + \underline{0}] = \beta
\end{aligned}
$$

$$
\begin{aligned}
Cov\,\widehat{\beta} &= Cov(AY) \quad \text{where } A = (X^t X)^{-1} X^t \\
&= A\,(\sigma_\epsilon^2\, I_n)\, A^t \\
&= \sigma_\epsilon^2\,(X^t X)^{-1} X^t \cdot X (X^t X)^{-1} \\
&= \sigma_\epsilon^2\,(X^t X)^{-1}
\end{aligned}
$$

Cute Example

$$Var(\widehat{\beta}_k) = Var\left([0 \cdots 0\,1\,0 \cdots 0] \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \vdots \\ \widehat{\beta}_k \\ \vdots \\ \widehat{\beta}_{p-1} \end{bmatrix} \right)$$

$$= Var(e_k^t AY)$$

$$= \sigma_\epsilon^2 e_k^t AA^t e_k$$

$$= \sigma_\epsilon^2\, e_k^t (X^t X)^{-1} e_k$$

$$= \sigma_\epsilon^2\, (X^t X)^{-1}_{kk}$$

as we have seen before.

Familiar Example $p = 1$

$$Y = X\beta + \epsilon \qquad \text{with} \qquad X = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

Thus

$$\hat{\beta} = \underbrace{(X^tX)^{-1}}_{n^{-1}} \underbrace{X^tY}_{n\bar{y}} = \bar{y}$$

$$\sigma^2(\hat{\beta}) = \sigma^2(\bar{y}) = \sigma^2_\epsilon (X^tX)^{-1} = \frac{\sigma^2_\epsilon}{n}$$

as usual.

Looking at the Criterion

$$\min_{\beta} \epsilon^t \epsilon = \sum_{i=1}^{n} (y_i - \beta)^2 \quad \text{(see sketch)}$$

Taylor's series:

$$g(\beta) = g(\widehat{\beta}) + (\beta - \widehat{\beta}) g'(\widehat{\beta}) + \frac{1}{2}(\beta - \widehat{\beta})^2 g''(\widehat{\beta}) + \cdots$$

$$g(\beta) = \sum (y_i - \beta)^2 \qquad g(\widehat{\beta}) = \sum (y_i - \bar{y})^2$$

$$\begin{aligned} g'(\beta) &= -2 \sum (y_i - \beta) \\ &= -2n\bar{y} + 2n\beta \end{aligned} \qquad g'(\widehat{\beta}) = 0$$

$$g''(\beta) = 2n \qquad g''(\widehat{\beta}) = 2n$$

Therefore, $\quad g(\beta) = g(\widehat{\beta}) + 0 + n(\beta - \widehat{\beta})^2 + \cdots$

Note:

$$\sum (y_i - \beta)^2 = \sum (y_i - \widehat{\beta} + \widehat{\beta} - \beta)^2$$
$$= \sum \left[(y_i - \widehat{\beta})^2 + 2(\widehat{\beta} - \beta)(y_i - \widehat{\beta}) + (\widehat{\beta} - \beta)^2 \right]$$
$$= \sum (y_i - \bar{y})^2 + 0 + n(\widehat{\beta} - \beta)^2 \text{ exactly!!}$$

Dual: $n = 100$ and $n = 400$ (see sketches)

Tentative conclusion:

steeper criterion $\Rightarrow$ more accurate parameters

Multivariate β

$$
\begin{aligned}
g(\beta) &= \epsilon^t \epsilon \\
Y &= X\beta + \epsilon \\
\widehat{\beta} &\sim N(\beta, \sigma_\epsilon^2 (X^t X)^{-1})
\end{aligned}
$$

Multivariate Taylor's series:

$$
g(\beta) = g(\widehat{\beta}) + (\beta - \widehat{\beta})^t \nabla g(\widehat{\beta}) + \frac{1}{2}(\beta - \widehat{\beta})^t \nabla^2 g(\widehat{\beta})(\beta - \widehat{\beta}) + \cdots
$$

$$g(\beta) = Y^t Y - 2\beta^t X^t Y + \beta^t X^t X \beta$$

$$\nabla g(\beta) = \underline{0} - 2X^t Y + 2X^t X \beta$$

Hessian is ∇ of $\nabla^t g(\beta)$; so

$$\nabla^t g(\beta) = -2Y^t X + 2\beta^t X^t X$$

and

$$\nabla \nabla^t g(\beta) = -\underline{0} + 2X^t X \,.$$

For our least-squares problem:

$$\begin{aligned} g(\beta) &= g(\widehat{\beta}) + 0 + \frac{1}{2}(\beta - \widehat{\beta})^t 2(X^t X)(\beta - \widehat{\beta}) \\ &= Y^t(I - H)Y + (\beta - \widehat{\beta})^t (X^t X)(\beta - \widehat{\beta}) \end{aligned}$$

(see sketch)

Facts about Positive Definite Matrices

$$A = X^t X \qquad \text{symmetric}$$

Look at the quadratic form

$$y^t A\, y = y^t X^t \underbrace{X\, y}_{w} = w^t w \geq 0 \quad \forall y$$

Look at eigenvalues/eigenvectors:

$$A\,v_k = \lambda_k\,v_k \qquad k = 1, \ldots, p$$

and

$$v_k^t v_k = 1 \qquad v_k^t v_\ell = 0 \quad k \neq \ell\,.$$

Assume $\lambda_1 > \lambda_2 > \cdots > \lambda_p$. Consider

$$v_k^t\,(A\,v_k) = v_k^t\,(\lambda_k\,v_k) = \lambda_k\,(v_k^t v_k) = \lambda_k > 0$$

so, in fact, all the $\lambda_k > 0$.

Definitions: A symmetric matrix

A is p.d. if all $\lambda_k > 0$

A is p.s.d. if all $\lambda_k \geq 0$

A is n.d. if all $\lambda_k < 0$

A is n.s.d. if all $\lambda_k \leq 0$

A is indefinite if some $\lambda_k > 0$ & some $\lambda_k < 0$

Level Sets

$$g(\beta) = g(\widehat{\beta}) + (\beta - \widehat{\beta})^t A \, (\beta - \widehat{\beta})$$

Find values of β satisfying

$$(\beta - \widehat{\beta})^t A \, (\beta - \widehat{\beta}) = c$$

(see sketch)

Suppose $$A = \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_p \end{pmatrix}$$

By inspection, $A\,e_k = a_k\,e_k$, so eigenvectors are the coordinate axes. Level sets satisfy

$$y^t\,A\,y = \sum_{k=1}^{p} a_k\,y_k^2 = \sum_{k=1}^{p} \frac{y_k^2}{1/a_k} = c$$

which is an ellipse.

(see sketch)

Note:

$$
\begin{aligned}
Av_k = \lambda_k v_k \quad &\Rightarrow \quad v_k = \lambda_k A^{-1} v_k \\
&\Rightarrow \quad \frac{1}{\lambda_k} v_k = A^{-1} v_k
\end{aligned}
$$

so the eigenvectors of A and A^{-1} are the *same*, while the eigenvalues are reciprocal of each other.

Next, find point, y, on the level set in the direction, v_1; thus y has the form $\alpha\, v_1$:

$$
\begin{aligned}
y^t\, A\, y &= c \\
\alpha^2\, \underbrace{v_1^t\, \underbrace{A\, v_1}_{\lambda_1\, v_1}}_{\lambda_1} &= c
\end{aligned}
$$

so that

$$\alpha^2 = \frac{c}{\lambda_1}$$

Since λ_1 is the largest eigenvalue, this is the shortest axis of the ellipse. Also, changes in β in that direction give the quickest increase in the criterion function, BUT most accurate in that direction.

In general, points $y_k = \alpha_k v_k$ on the level set satisfy

$$\alpha_k^2 v_k^t A v_k = c \quad \Rightarrow \quad \alpha_k = \sqrt{\frac{c}{\lambda_k}}$$

$$y_k = \pm\sqrt{\frac{c}{\lambda_k}} v_k .$$

In $\Re^2$, see sketch of $g(\beta)$ about $g(\widehat{\beta})$...

Recall that $\sigma^2(\widehat{\beta}) = \sigma_\epsilon^2\,(X^t X)^{-1}$. Thus,

$$Var(w^t\widehat{\beta}) = w^t\,\sigma_\epsilon^2\,(X^t X)^{-1}\,w$$

Look in the direction $w = v_k$:

$$\sigma_\epsilon^2\,v_k^t\,(X^t X)^{-1}\,v_k = \frac{\sigma_\epsilon^2}{\lambda_k}$$

therefore

$$std(w^t\beta) = \frac{\sigma_\epsilon}{\sqrt{\lambda_k}}$$

see sketch...

or get same result by recalling

$$\beta - \widehat{\beta} \sim N(0, \sigma_\epsilon^2 (X^t X)^{-1})$$

has level sets

$$(\beta - \widehat{\beta})^t \Sigma^{-1} (\beta - \widehat{\beta}) = c$$

Look in the direction $\beta - \widehat{\beta} = \alpha_k v_k$, then

$$\begin{aligned} \alpha_k^2 v_k^t \left(\sigma_\epsilon^2 (X^t X)^{-1} \right)^{-1} v_k &= c \\ \frac{\alpha_k^2}{\sigma_\epsilon^2} \underbrace{v_k^t (X^t X) v_k}_{\lambda_k} &= c \end{aligned}$$

so

$$\alpha_k^2 = \frac{\sigma_\epsilon^2 \, c}{\lambda_k} \quad \Rightarrow \quad \alpha_k = \sigma_\epsilon \sqrt{\frac{c}{\lambda_k}}$$

(see sketch).... note the same orientation in the end.

THE END

Well, now for the computer demos...