smfl13.tex

Lecture 13. 14.2.2011

Matrices of the form $A^T A$ are common in Statistics (e.g., in Regression).
1. $A^T A$ is always non-negative definite, since $x^T A^T A x = (Ax)^T (Ax) = y^T y = \sum y_i^2 \geq 0$, with $y := Ax$. So all eigenvalues of $A^T A$ are non-negative.
2. $A^T A$ is positive definite iff all eigenvalues are positive iff $A^T A$ is non-singular, and one can show this happens iff A has full rank.
3. If $N(A)$ is the null space of A (the vector space of all x with $Ax = 0$), $N(A) = N(A^T A)$.
4. $A^T A$ and A have the same rank.

2. SINGULAR VALUE DECOMPOSITION (SVD).

The following algebraic result is extremely important in Statistics, and in Numerical Analysis. I do not have a reference in a standard Algebra or Linear Algebra book; I have used [HJ] 3.0, 3.1, [GvL] 2.5.

Theorem (Singular Value Decomposition, SVD). If A $(n \times p)$ has rank r, A can be written

$$A = ULV^T,$$

where U $(n \times r)$ and V $(p \times r)$ are column-orthogonal ($U^T U = V^T V = I_r$) and L $(r \times r)$ is a diagonal matrix with positive elements, and

$$A = \sum_{i=1}^{r} \lambda_i u_i v_i^T,$$

where
(i) the λ_i are the square roots of the positive eigenvalues of $A^T A$ (or AA^T) – the *singular values*;
(ii) the vectors u_i, v_i are eigenvectors of AA^T and $A^T A$ – the *left* and *right singular vectors*.
(For A square and symmetric, this reduces to the Spectral Decomposition).

Proof. Since A has rank r, so by above does $A^T A$. Apply Spectral Decomposition to $A^T A$, obtaining the $\Lambda = diag(\lambda_i)$, Γ as above. The λ_i are non-negative (as $A^T A$ is non-negative definite). The $\lambda_i = 0$ terms make no contribution to the sum $\sum \lambda_i \gamma_i \gamma_i^T$, so we omit the zero eigenvalues and their eigenvectors. Write the result as

$$A^T A = V \Lambda V^T,$$

where $\Lambda = diag(\lambda_i)$ contains the positive eigenvalues and V $(p \times r)$ is a column-orthogonal matrix of their eigenvectors. Write

$$\ell_i := \lambda_i^{1/2} \quad (i = 1, \ldots, r), \quad L := diag(\ell_1, \ldots, \ell_r).$$

Define U $(n \times r)$ by

$$u_i := \ell_i^{-1} A v_i \qquad (i = 1, \ldots, r),$$

and write $U = (u_1, \ldots, u_r)$, $V = (v_1, \ldots, v_r)$ as rows of column-vectors. Then

$$u_j^T u_i = \ell_i^{-1} \ell_j^{-1} v_j^T A^T A v_i = \lambda_i \ell_i^{-1} \ell_j^{-1} v_j^T v_i,$$

as v_i an eigenvector of $A^T A$ corr. to λ_i. The RHS is δ_{ij} by orthogonality of the u_i, and $\ell_i = \lambda_i^{1/2}$. So U is also column-orthogonal.

Since $A^T A$ $(p \times p)$ has rank r, and r orthogonal eigenvectors $v_1, \ldots, v_r$, which span its column-space $C(A^T A)$, any p-vector x can be written as

$$x = \sum_1^r \alpha_i v_i + y,$$

with $y \in N(A^T A)$, or $y \in N(A)$ by above. Now $N(A^T A)$ is the eigenspace of $A^T A$ for the eigenvalue 0, so y is orthogonal to the eigenvectors v_i of $A^T A$ corresponding to the non-zero eigenvalues. Let e_i be the column-vector of length n with 1 in the ith place, 0 elsewhere. Then

$$ULV^T x = UL \begin{pmatrix} v_1^T \\ \vdots \\ v_r^T \end{pmatrix} \left(\sum \alpha_i v_i + y\right) = UL \begin{pmatrix} \sum \alpha_i v_1^T v_i + v_1^T y \\ \vdots \\ \sum \sigma_i v_r^T v_i + v_r^T y \end{pmatrix}$$

$$= UL \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_r \end{pmatrix} = \sum \alpha_i U L e_i.$$

As

$$ULe_i = (u_1, \ldots, u_r) \begin{pmatrix} \ell_1 & & \\ & \ddots & \\ & & \ell_r \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = (u_1 \ell_1, \ldots, u_r \ell_r) \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = u_i \ell_i = \ell_i u_i,$$

this gives

$$\begin{aligned} ULV^T x &= \sum \alpha_i \ell_i u_i \\ &= \sum \alpha_i A v_i \qquad (\ell_i u_i = A v_i) \\ &= \sum \alpha_i A v_i + Ay \qquad (Ay = 0 \text{ as } y \in N(A)) \\ &= A(\sum \alpha_i v_i + y) \\ &= Ax. \end{aligned}$$

Since this holds for all x,

$$ULV^T = A. \qquad //$$

Generalised Inverses and SVD.

Recall that the generalised inverse A^- of A satisfies $AA^-A = A$. If A has SVD $A = ULV^T$, one can check that

$$A^- := VL^{-1}U^T$$

is a generalised inverse of A.

3. STATISTICAL SETTING.

Usually in Statistics we have *univariate data* $x = (x_1, \ldots, x_n)$, and have to analyse it. Sometimes, however, each observation contains several different readings (measurements, for example) on the same 'individual', or object. We then need a two-suffix notation just to describe the data, and so we use matrices throughout.

Notation. We assume that p variables are measured on each of n objects. We assemble the np readings into a *data matrix*

$$X = \begin{pmatrix} x_{11} & \ldots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \ldots & x_{np} \end{pmatrix},$$

where x_{ij} is the observation on the jth variable measured on the ith reading.

As always, n may be large – the larger the better, as large samples are more informative than small ones. The size of p varies with the problem. But typically p might be of the order of 10 or 12, say. A 12-dimensional 'variable space' is unwieldy for many purposes, and we might want a lower-dimensional representation of the data, with as little loss of information as

possible. Background: [MKB] Ch. 1, [K] Ch. 1.
Notation.

$$X = (x_{(1)}, \ldots, x_{(p)}) = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix}.$$

So the *column-vectors* x_i, $x_{(j)}$ relate to the ith *object* and the jth *variable*.
Mean vector. $\overline{x}_i := \frac{1}{n}\sum_{r=1}^n x_r i$ is the sample mean of the ith variable; the *sample mean vector* is

$$\overline{x} := \begin{pmatrix} \overline{x}_1 \\ \vdots \\ \overline{x}_p \end{pmatrix}.$$

The sample variance s_{ij} between the ith and jth variables is

$$s_{ij} := \frac{1}{n}\sum_{r=1}^{n}(x_{ri} - \overline{x}_i)(x_{rj} - \overline{x}_j) = \frac{1}{n}\sum_{r=1}^{n} x_{ri}x_{rj} - \overline{x}_i\overline{x}_j.$$

Form these into a matrix, the *sample covariance matrix* $S := (s_{ij})$:

$$S = \frac{1}{n}\sum_{r=1}^{n}(x_r - \overline{x})(x_r - \overline{x})^T = \frac{1}{n}\sum_{r=1}^{n} x_r x_r^T - \overline{x}\;\overline{x}^T.$$

Now $X^T = (x_1, \ldots, x_n)$ (row of columns corresponding to *objects*), so

$$XX^T = (x_1, \ldots, x_n)\begin{pmatrix} x_1^T \\ \vdots \\ x_n \end{pmatrix} = \sum x_r x_r^T.$$

Write $\mathbf{1}$ for a column-vector of n 1s. Then (check) $\mathbf{1}\mathbf{1}^T$ is the $n \times n$ matrix with each element 1, and (check) $X^T\mathbf{1}\mathbf{1}^T X = n^2\overline{x}\;\overline{x}^T$. So

$$S = \frac{1}{n}X^T X - \frac{1}{n^2}X^T\mathbf{1}\mathbf{1}^T X = \frac{1}{n}X^T H X, \quad \text{where} \quad H := I - \frac{1}{n}\mathbf{1}\mathbf{1}^T$$

is the $n \times n$ *centring matrix*. We call $M := X^T X = \sum_1^n x_r x_r^T$ the *matrix of sums of squares and products.*