ACM 104

Homework Set 4 Solutions

February 14, 2001

1 Franklin Chapter 2, Problem 4, page 55.
Suppose that we feel that some observations are more important or reliable than others. Redefine the function to be minimized as:

$$\phi(x) = \sum_{i=1}^{m} \sigma_i^{-2}(b_i - x_0 - x_i a_{i1} - x_2 a_{i2} - \ldots - x_n a_{in})^2$$

Now what are the appropriate normal equations? Assume real data.

Following the procedure of p.51, we differentiate the above relation with respect to x_j:

$$0 = -\frac{1}{2}\frac{\partial \phi}{\partial x_j} = \sum_{i=1}^{m} a_{ij}\sigma_i^{-2}(b_i - x_0 - x_i a_{i1} - x_2 a_{i2} - \ldots - x_n a_{in})$$

Now define

$$b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}, A = \begin{pmatrix} 1 & a_{11} & \ldots & a_{1n} \\ 1 & a_{21} & \ldots & a_{2n} \\ \cdot & \cdot & \ldots & \cdot \\ 1 & a_{m1} & \ldots & a_{mn} \end{pmatrix}, x = \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_1^{-2} & 0 & \ldots & 0 \\ 0 & \sigma_2^{-2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_m^{-2} \end{pmatrix}$$

The normal equations can be written as:

$$0 = A^T\Sigma(b - Ax) \Rightarrow A^T\Sigma A x = A^T\Sigma b$$

2 Franklin Chapter 4, Problem 3, page 75.
Let

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\alpha_4 & -\alpha_3 & -\alpha_2 & -\alpha_1 \end{pmatrix}, x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$$

Can x be an eigenvector of this matrix, A, if $x_1 = 0$? If λ is an eigenvalue of A, and if $x_1 = 1$, what is the form of the eigenvector x? What is the characteristic polynomial of A? Under what condition does A have four linearly independent eigenvectors?

If x were an eigenvector of A with $x_1 = 0$ then there would exist $\gamma \neq 0$ such that

$$= \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\alpha_4 & -\alpha_3 & -\alpha_2 & -\alpha_1 \end{pmatrix} \begin{pmatrix} 0 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \gamma \begin{pmatrix} 0 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$$

i.e. $x_2 = 0, x_3 = \gamma x_2, x_4 = \gamma x_3, -(\alpha_3 x_2 + \alpha_2 x_3 + \alpha_1 x_4) = \gamma x_4$. This implies $x_2 = x_3 = x_4 = 0$ which means that x is not an eigenvector since it is not a nonzero vector. We conclude that x cannot be an eigenvector if $x_1 = 0$. Now let λ be an eigenvalue and $x_1 = 1$. Then we have

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\alpha_4 & -\alpha_3 & -\alpha_2 & -\alpha_1 \end{pmatrix} \begin{pmatrix} 1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$$

i.e. $x_2 = \lambda, x_3 = \lambda x_2, x_4 = \lambda x_3, -(\alpha_4 + \alpha_3 x_2 + \alpha_2 x_3 + \alpha_1 x_4) = \lambda x_4$. So the form of the eigenvector is

$$x = \begin{pmatrix} 1 \\ \lambda \\ \lambda^2 \\ \lambda^3 \end{pmatrix}$$

with $\alpha_4 + \lambda\alpha_3 + \lambda^2\alpha_2 + \lambda^3\alpha_1 + \lambda^4 = 0$. The characteristic polynomial of a matrix is a polynomial of degree four, where the coefficient of the highest power in λ is 1, whose roots are the eigenvalues of the matrix. Looking at the previous relation for the eigenvalues we realize that

$$p(\lambda) = \lambda^4 + \alpha_1\lambda^3 + \alpha_2\lambda^2 + \alpha_3\lambda + \alpha_4$$

is the characteristic polynomial. If the characteristic polynomial has 4 distinct nonzero roots then there will be 4 linearly independent eigenvectors. This is a sufficient condition for 4 linearly independent eigenvectors but it may not be a necessary conditon since, in general, distinct eigenvectors can be obtained from the same eigenvalues. However, in this case we can show that this is a necessary condition. We can always scale the x_1 component to be 1. Suppose that we have the following 4 eigenvectors

$$\begin{pmatrix} 1 \\ \lambda_1 \\ \lambda_1^2 \\ \lambda_1^3 \end{pmatrix}, \begin{pmatrix} 1 \\ \lambda_2 \\ \lambda_2^2 \\ \lambda_2^3 \end{pmatrix}, \begin{pmatrix} 1 \\ \lambda_3 \\ \lambda_3^2 \\ \lambda_3^3 \end{pmatrix}, \begin{pmatrix} 1 \\ \lambda_4 \\ \lambda_4^2 \\ \lambda_4^3 \end{pmatrix}$$

These are linearly independent if and only if the determinant of the vectors (as rows or columns) is not zero. So consider

$$\begin{vmatrix} 1 & \lambda_1 & \lambda_1^2 & \lambda_1^3 \\ 1 & \lambda_2 & \lambda_2^2 & \lambda_2^3 \\ 1 & \lambda_3 & \lambda_3^2 & \lambda_3^3 \\ 1 & \lambda_4 & \lambda_4^2 & \lambda_4^3 \end{vmatrix} = \prod_{i>j} (\lambda_i - \lambda_j)$$

(result of Vandermonde matrices). This is nonzero if and only if $\lambda_i \neq \lambda_j$ for $i \neq j$. Therefore, we need to have distinct eigenvalues for this matrix in order to get distinct eigenvectors.

3 Franklin Chapter 4, Problem 4, page 75.
Generalize the last problem, replacing 4 by n. (The matrix, A, is known as the companion matrix of the polynomial $\lambda^n + \alpha_1\lambda^{n-1} + \ldots + \alpha_n$.

If x were an eigenvector of A with $x_1 = 0$ then there would exist $\gamma \neq 0$ such that

$$\begin{pmatrix} 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \\ -\alpha_n & -\alpha_{n-1} & \ldots & -\alpha_1 \end{pmatrix} \begin{pmatrix} 0 \\ x_2 \\ \cdot \\ x_n \end{pmatrix} = \gamma \begin{pmatrix} 0 \\ x_2 \\ \cdot \\ x_n \end{pmatrix}$$

which implies, as before that $(x_1, \ldots, x_n)$ is the zero vector, so there are no eigenvectors of the companion matrix with $x_1 = 0$. Let λ be an eigenvalue and $x_1 = 1$. Then we have

$$\begin{pmatrix} 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \\ -\alpha_n & -\alpha_{n-1} & \ldots & -\alpha_1 \end{pmatrix} \begin{pmatrix} 1 \\ x_2 \\ \cdot \\ x_n \end{pmatrix} = \gamma \begin{pmatrix} 1 \\ x_2 \\ \cdot \\ x_n \end{pmatrix}$$

So the eigenvector is

$$x = \begin{pmatrix} 1 \\ \lambda \\ \lambda^2 \\ \vdots \\ \lambda^n \end{pmatrix}$$

with $\alpha_n + \lambda\alpha_{n-1} + \ldots + \lambda^{n-1}\alpha_1 + \lambda^n = 0$. The characteristic polynomial of an $n \times n$ matrix is a polynomial of degree n, where the coefficient of the highest power in λ is 1, whose roots are the eigenvalues of the matrix. Looking at the previous relation for the eigenvalues we realize that

$$p(\lambda) = \alpha_n + \lambda\alpha_{n-1} + \ldots + \lambda^{n-1}\alpha_1 + \lambda^n$$

is the characteristic polynomial. If the characteristic polynomial has n distinct nonzero roots then there will be n linearly independent eigenvectors. We will show that this is also a necessary condition. We can always scale the x_1 component to be 1. Suppose that we have the n eigenvectors which arise from the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ They are linearly independent if and only if the determinant of the matrix that they can form (as rows or columns) is not zero. So consider

$$\begin{vmatrix} 1 & \lambda_1 & \lambda_1^2 & \ldots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \ldots & \lambda_2^{n-1} \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ 1 & \lambda_n & \lambda_n^2 & \ldots & \lambda_n^{n-1} \end{vmatrix} = \prod_{i>j} (\lambda_i - \lambda_j)$$

(result of Vandermonde matrices). This is nonzero if and only if $\lambda_i \neq \lambda_j$ for $i \neq j$. Therefore, we need to have distinct eigenvalues for this matrix in order to get distinct eigenvectors.

4 Franklin Chapter 4, Problem 5, page 88.
Assume that $\rho_1 > 0, \ldots, \rho_n > 0$. Redefine the "length" of a vector as

$$|x| \equiv (\rho_1 |x_1|^2 + \ldots + \rho_n |x_n|^2)^{\frac{1}{2}}$$

Which matrices V preserve this "length"?

Let

$$Q = \begin{pmatrix} \rho_1^{\frac{1}{2}} & & 0 \\ & \ddots & \\ 0 & & \rho_n^{\frac{1}{2}} \end{pmatrix}$$

so that

$$Q^2 = P = \begin{pmatrix} \rho_1 & & 0 \\ & \ddots & \\ 0 & & \rho_n \end{pmatrix}$$

then we can write the new "length" as

$$|x|_P = (Qx, Qx)^{\frac{1}{2}} = (Px, x)^{\frac{1}{2}}$$

or $|x|_P^2 = (Px, x)$. We will first show that $|Vx|_P = |x|_P$ if and only if $(QVx, QVy) = (Qx, Qy)$. For the only if part, suppose that $|Vx|_P = |x|_P$. Then we can write

$$|x+y|_P^2 = \sum \rho_i |x_i + y_i|^2 = \sum \rho_i |x_i|^2 + \sum \rho_i x_i \overline{y}_i + \sum \rho_i \overline{x}_i y_i + \sum \rho_i |y_i|^2$$

i.e. we have

$$|x+y|_P^2 = |x|_P^2 + 2\mathrm{Re}(Px, y) + |y|_P^2 \tag{1}$$

Replacing $x+y$ with $x+iy$ we get

$$|x+iy|_P^2 = |x|_P^2 + 2\mathrm{Im}(Px, y) + |y|_P^2 \tag{2}$$

Now replacing x with Vx and y with Vy in (1) and (2) we obtain

$$|V(x+y)|_P^2 = |Vx|_P^2 + 2\mathrm{Re}(PVx, Vy) + |Vy|_P^2 \tag{3}$$

$$|V(x+iy)|_P^2 = |Vx|_P^2 + 2\mathrm{Im}(PVx, Vy) + |Vy|_P^2 \tag{4}$$

Since we assumed that $|Vx|_P = |x|_P$ we must have after equating (1) with (3) and (2) with (4)

$$\mathrm{Re}(PVx, Vy) = \mathrm{Re}(Px, y),\ \mathrm{Im}(PVx, Vy) = \mathrm{Im}(Px, y)$$

so $(PVx, Vy) = (Px, y)$ implies that $(V^*PVx, y) = (Px, y)$, i.e. $(QVx, QVy) = (Qx, Qy)$. Now assuming that $(QVx, QVy) = (Qx, Qy)$ we have that setting $x = y$

$$(QVx, QVx) = (Qx, Qx) \Rightarrow |Vx|_P^2 = |x|_P^2$$

which implies that $|Vx|_P = |x|_P$. Now taking the result $(V^*PVx, y) = (Px, y)$ which is true for all x, y we set $x = e^i$, $y = e^j$ to pick out certain rows and columns. Then $(V^*PV_{ji} = P_{ij}$ so $V^*PV = P$ since components are equal. Therefore, if we have $V^*PV = P$ then V preserves

$$|x| \equiv (\rho_1 |x_1|^2 + \ldots + \rho_n |x_n|^2)^{\frac{1}{2}}$$

5 Franklin Chapter 4, Problem 4, page 94.
Let $a^1, \ldots, a^m$ be linearly independent vectors over the field of complex numbers. Define $\alpha_{ij} = (a^i)^* a^j = (a^j, a^i)$. Show that $\det(\alpha_{ij}) \neq 0$ by showing that an eigenvector x of the matrix (α_{ij}) belonging to the eigenvalue 0, would satisfy the relation $\| \sum x_i a^i \|^2 = 0$.

If we define

$$A = \begin{pmatrix} \vdots & \vdots & & \vdots \\ a^1 & a^2 & \ldots & a^m \\ \vdots & \vdots & & \vdots \end{pmatrix}$$

The matrix (α_{ij}) can be written as A^*A, where by A^* we mean the complex conjugate of the transpose of A.
If zero is an eigenvalue of (α_{ij}) with eigenvector x we then have

$$A^*Ax = 0$$

Then

$$0 = \langle x, A^*Ax \rangle = \langle Ax, Ax \rangle = \|Ax\|^2 \Rightarrow x = 0$$

since A has independent columns.

6 Let B be an $m \times n$ matrix and A be a square invertible matrix (mxm). Prove that AB and B have the same nullspace and the same row space (and the same rank) as B itself. Use this fact to show that the rank of a matrix is an invariant.

For $x \in R^n$ we have

$$Bx = 0 \Rightarrow ABx = 0$$

so $N(B) \subseteq N(AB)$. Also, since A is invertible,

$$ABx = 0 \Rightarrow A^{-1}ABx = A^{-1}0 \Rightarrow Bx = 0$$

so $N(AB) \subseteq N(B)$. These two relations imply $N(B) = N(AB)$.

The row space of a matrix is the space of all row vectors that can be written as a linear combination of the row vectors of that matrix:

$$\text{row space}(B) = \left\{\text{all vectors in the form } x^T B, \text{for some } x^T \in R^m\right\}$$

$$\text{row space}(AB) = \left\{\text{all vectors in the form } y^T AB, \text{for some } y^T \in R^m\right\}$$

$y^T A$ is a vector in R^m so every element in the row space of AB is also in the row space of B. Any element $x^T B$ in the row space of B is also in the row space of AB and the corresponding y^T is $y^T = x^T A^{-1}$.Then the row spaces of A and AB must be the same, and the matrices have the same rank.

Let $X \sim Y$ denote the fact that two matrices X and Y have the same rank. Applying what we proved to the case $m = n$ we have that when B is an $n \times n$ matrix and A is an $n \times n$ invertible matrix then $B \sim AB$. Since taking the transpose of an $n \times n$ matrix does not change its rank, $B \sim B^T \sim AB^T \sim BA^T$. This implies that $B \sim BA$, when A is invertible.
Now let A and C be the same up to a coordinate transformation, that is let $A = TCT^{-1}$, for some invertible matrix T. Then $C \sim TC \sim TCT^{-1} \sim A$ and the rank of a matrix is invariant.

7 Let A and B be two square matrices. Prove that the characteristic polynomials of BA and AB are identical. (Hint: It may be helpful first to establish the result for $\det A \neq 0$. In the case where $\det A = 0$, think about 'perturbing' the matrix A such that it becomes invertible.)

First suppose $\det(A) \neq 0$. Then A is invertible and:

$$\det(AB - \lambda I) = \det(A^{-1})\det(AB - \lambda I)\det(A) = \det(BA - \lambda I)$$

The matrices AB and BA have the same eigenvalues and therefore the same characteristic polynomial.

If $\det(A) = 0$ we know from problem set 3 exercise 4 that for $\epsilon > 0$ small enough,$\det(A + \epsilon I) \neq 0$. Applying what we just showed to the matrices $A + \epsilon I$, B we conclude that

$$AB + \epsilon B \text{ and } BA + \epsilon B \text{ have the same eigenvalues } \lambda(\epsilon), \text{ for every } \epsilon > 0 \text{ that is small enough} \quad (5)$$

For a matrix A, the coefficients of the polynomial $\det(A - \lambda I)$ depend continuously on the coefficients of A. $\det(A - \lambda I) = \prod(\lambda - \lambda_i)$ implies that the coefficients of the characteristic polynomial depend continuously on the eigenvalues (actually, the eigenvalues are a polynomial of the coefficients). Since the eigenvalues of a matrix depend continuously on the matrix, (5) that must hold also for $\epsilon = 0$.

8 A projection matrix is a matrix which satisfies the identity $P = P^2$.

1. If $P = P^T P$ show that P is a projection matrix.
2. If P is an orthogonal projection matrix, show that $P = P^T$.
3. Suppose that P is the projection onto the subspace M and Q is the projection onto the orthogonal complement $M^\perp$. What are $P+Q$ and PQ? Show that $P-Q$ is its own inverse.

1. Taking the transpose of both sides in $P = P^T P$, we get $P^T = P^T P$, so $P = P^T$ and $P = P^T P = P^2$.
2. As we did for problem 5 in problem set 3, we can represent the orthogonal projection in a basis consisting of the r vectors that span the space onto which P projects and the $n-r$ vectors that are orthogonal to that space. In that basis, the matrix P is diagonal, with the first r elements being 1 and the rest 0. Let Λ be this diagonal matrix and let S be the orthogonal matrix whose columns are these basis vectors. We have

$$P = S\Lambda S^T \Rightarrow P = S\Lambda^T S^T = S\Lambda S^T = P.$$

   which proves the claim.

3. The orthogonal complement $M^\perp$ consists of vectors that are orthogonal to all vectors in M. The problem is again simplified if we choose as basis the $r = dim(M)$ eigenvectors of P with eigenvalue 1 together with the $n-r$ eigenvectors of Q with eigenvalue 1. Then the matrices that represent these projections are:

$$P = \begin{pmatrix} I_1 & 0 \\ 0 & 0 \end{pmatrix}, Q = \begin{pmatrix} 0 & 0 \\ 0 & I_2 \end{pmatrix}, \text{ with } I_1, I_2 r \text{ by } r \text{ and } (n-r) \text{ by } (n-r) \text{ identities}$$

   Then $P+Q$ is the identity in this basis and an $n \times n$ invertible matrix in general, PQ is zero and $P-Q$ is indeed its own inverse.
   To express any matrix in a different basis we multiply on the left by T and on the right by T^{-1}. By inspection we see that these results hold, even when P and Q are expressed in an arbitrary basis.

   Inuitively they really make sense: $Px + Qx = x$ means that the sum of the components of x in M and $M^\perp$ give back the vector, $PQx = 0$ means if we try to project $Qx \in M^\perp$ onto M we will get the zero vector, and using these two results:

$$(P-Q)(P-Q) = P^2 - PQ - QP + Q^2 = P^2 + Q^2 = P + Q = I$$