NAME: Nir Edelman !!

CS170A — Mathematical Models and Methods for Computer Science

Spring 2011

D.S. Parker, Akshay Wadia

Midterm Examination

OPEN BOOK, OPEN NOTES

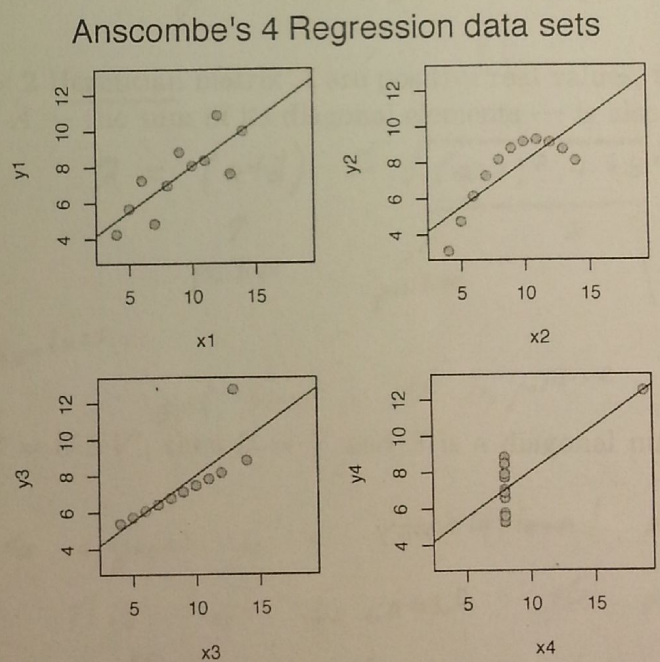
Wednesday, May 4, 4:00pm–5:50pm

Do not cheat.

Problem	*Points*
1	21 /25
2	18 /25
3	25 /25
4	19 /25
Total	83 /100

Throughout this exam we will make use of the *Anscombe datasets* — four small $\boldsymbol{x}, \boldsymbol{y}$ datasets with some identical basic statistics but very different shape:

	x_1	x_2	x_3	x_4	y_1	y_2	y_3	y_4
1	10	10	10	8	8.04	9.14	7.46	6.58
2	8	8	8	8	6.95	8.14	6.77	5.76
3	13	13	13	8	7.58	8.74	12.74	7.71
4	9	9	9	8	8.81	8.77	7.11	8.84
5	11	11	11	8	8.33	9.26	7.81	8.47
6	14	14	14	8	9.96	8.10	8.84	7.04
7	6	6	6	8	7.24	6.13	6.08	5.25
8	4	4	4	19	4.26	3.10	5.39	12.50
9	12	12	12	8	10.84	9.13	8.15	5.56
10	7	7	7	8	4.82	7.26	6.42	7.91
11	5	5	5	8	5.68	4.74	5.73	6.89



Each of the four $\boldsymbol{x}$ vectors has the same mean value (9.0) and variance (5.5).
Each of the four $\boldsymbol{y}$ vectors has the same mean value (7.5) and variance (4.12).
Furthermore, for each of the four values of i, the $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ have the same covariance (5.5).

These datasets also share one least squares fit $y = ax + b$, where $a = 1/2$ and $b = 3$. That is, the coefficients a and b are the same for each of the datasets.

Define the 11×2 matrix D_i as the matrix that has the two columns $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$, for $1 \leq i \leq 4$.

1. SVD

For each of the following matrices, determine (yes/no) whether the matrix is: unitary, hermitian, invertible, positive definite.

matrix	unitary?	hermitian?	invertible?	positive definite?
$\frac{1}{\sqrt{2}}\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$	True ☐ False ■	True ■ False ☐	True ☐ False ■	True ☐ False ■
$\frac{1}{\sqrt{2}}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$	True ☐ False ■	True ■ False ☐	True ■ False ☐	True ☐ False ■
$\frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	True ☐ False ■	True ■ False ☐	True ■ False ☐	True ☐ False ■

For this examination, assume:

Definition: A is *positive definite* if it is Hermitian and all its eigenvalues are positive real values.

For each of the following equations, mark whether the equation is True (valid for all specified matrices A) or False. Assume that A' denotes the hermitian transpose of A, and $i = \sqrt{-1}$.

For full credit, if you mark it True, you must *explain* how you derived this. If you mark it False, you must give a *counterexample*.

(a) **True** ☐ **False** ■

If a 2×2 matrix A is Hermitian and has positive determinant, then A is positive definite.

Explanation: $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ $\lambda = \frac{(-2) \pm \sqrt{(0) + 0}}{2}$

eigenvalues not all positive

(b) **True** ■ **False** ☐

If both the determinant and trace of a 2×2 Hermitian matrix A are positive real values, then A is positive definite. (Hint: the trace of A — the sum of its diagonal elements — is also the sum of its eigenvalues.) -1

$\lambda = \frac{(a+d) \pm \sqrt{(a-d)^2 + 4b^2}}{2}$

Explanation: $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ positive positive

This will yield positive eigenvalues

and since det is positive, a+d will be greater than second part

Why?

(c) **True** ■ **False** ☐ -3

If C is a correlation matrix with SVD $C = U\,S\,V'$, then $U = V$ and S is a diagonal matrix with positive values on the diagonal.

Explanation: The values on the diagonal in a correlational matrix are all ~~~~ 1's. This is because the vector is compared with itself and the correlation is one

(d) **True** ■ **False** ☐

For any matrix A, the singular values of A are the square roots of the eigenvalues of $(A'\,A)$.

Explanation:

$A = USV'$ $A' = VS'U'$

$VS'U'USV' = VS'SV'$ since U is unitary

So: $S'S$ will give the square of all the values of S

Covariance

The discussion of the Anscombe data mentioned earlier defines four matrices, for $1 \le i \le 4$: D_i is the 11×2 matrix with the two columns x_i and y_i.

(a) Using the information about the Anscombe datasets above (do NOT calculate numerically), what differences are there between the four covariance matrices $cov(D_i)$, $1 \le i \le 4$?

The differences between the four covariance matrices are

the values of the vectors in the covariance matrix ?? ; but

the total covariance will be the same.

OK

(b) Suppose that we changed the vector x_1 by subtracting its mean 9 from each of its entries. How would $\mathrm{cov}(D_1)$ change, if at all?

$\sigma_{xy} = \frac{1}{n-1} \left((x - \mu_x) - \mu_x \right)' (y - \mu_y)$

since X_1 is subtracted by the same value, there will be overall no change in the covariance (D_1). This is because the same value is subtracted from x_1

(c) Suppose that we changed the vector x_1 by multiplying each of its entries by $1/\sigma$, where σ is the standard deviation of x_1. How would $\mathrm{cov}(D_1)$ change, if at all?

The covariance will change by a factor of $\left(\frac{1}{\sigma}\right)$

This is because $\sigma_{xy} = \frac{1}{n-1} \left(\frac{1}{\sigma_1} x_1 - \mu_x \right)(y_i - \mu_y)$

So all values will be scaled by $\left(\frac{1}{\sigma}\right)$. how?

$\begin{pmatrix} \sigma_x^2/\sigma^2 & \sigma_{xy}/\sigma \\ \sigma_{xy}/\sigma & \sigma_y^2 \end{pmatrix}$

−2

(d) Suppose that we changed y_1 to be x_1. How would $\mathrm{cov}(D_1)$ change, if at all?

$\mathrm{Cov}(D_1)$ would not change since $\sigma_{xy} = \frac{1}{n-1} \langle (x - \mu_x), (y - \mu_y) \rangle$

interchanging $y \leftrightarrow x$ reverses both rows and columns.

and this is equal to $\sigma_{yx} = \frac{1}{n-1} \langle (y - \mu_y), (x - \mu_x) \rangle$ $\begin{pmatrix} \sigma_y^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_x^2 \end{pmatrix}$

−2

(e) The fourth plot (for D_4) shows a single outlier at the eighth point (19, 12.5). If this (x, y)-pair were removed, the variance of the remaining 10 y_4 values would be 1.53. For the resulting 10×2 matrix D_4 then, what would $cov(D_4)$ be?

$\mathrm{cov}(D_4)$ will be lessened ? since there will be less variance between X and y

$\sigma_x^2 = \sigma_{xy} = 0.$

$\mathrm{cov}(D_4) = \begin{pmatrix} 0 & 0 \\ 0 & 1.53 \end{pmatrix}$

−3

3. **Least Squares**

For each of the Anscombe datasets ($1 \le i \le 4$), the least squares fit is the same: $\boldsymbol{y}_i = \boldsymbol{x}_i/2 + 3$.

Let '**1**' be the 11×1 vector of 1's, and let X_i be the matrix with the two columns $\boldsymbol{x}_i$ and **1**.

(a) For $1 \le i \le 4$, suppose we use least squares to solve

$X_i = (x_i \ 1)$

$$X_i \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \boldsymbol{y}_i.$$

What must be the resulting values for α_i and β_i ($1 \le i \le 4$)?

$x_i \alpha_i + \beta_i = y_i$

$\alpha_i = \frac{1}{2} \qquad \beta_i = 3$

(b) Show that $\|\boldsymbol{y}_i\|_2^2 = (n-1)\,(\,\mathrm{var}(\boldsymbol{y}_i) - n\ \mathrm{mean}(\boldsymbol{y}_i)^2\,)$, where $n = 11$, and therefore that the norms of all four vectors $\boldsymbol{y}_i$ are the same.

Also: $\sigma_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \mu_y)^2$

$\boxed{\|y_i\|_2^2 = \sum_{i=1}^{n} y_i^2}$

Therefore: $\boxed{\sum_{i=1}^{n} y_i^2 = (n-1)\sigma_y^2 + 2 y_i n \mu_y - n \mu_y^2}$

$(n-1)\sigma_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \mu_y)^2 = \sum_{j=1}^{n} y_i^2 - 2y_i\mu_y + \mu_y^2 = \sum_{i=1}^{n} (y_i^2 - 2y_i\mu_y) + n\mu_y^2$

(c) Explain why the R^2 values will be the same for the results of least squares on the first three datasets (using X_i and $\boldsymbol{y}_i$, $1 \le i \le 3$). (Hint: consider the results of the previous question.)

$R^2 = \frac{X X^- y_i}{\|y_i\|^2}$

The y_i's for the first 3 data sets have the same variance and mean and the same n

Therefore, the R^2 value will be the same

(d) Is it necessary that all of the pseudoinverses X_i^- be the same, for $1 \le i \le 4$? Explain why or why not.

$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = X_i^- y_i$

No, because although α_i, β_i are the same for $1 \le i \le 4$, y_i is not the same. Therefore X_i^- do not all have to be the same.

(e) Give a matrix formula for finding coefficients of a quadratic polynomial fitting the dataset D_2 using least squares.

$b = Ax \quad \Rightarrow \quad x = A^- b$

$b = \begin{pmatrix} y_2(1) \\ \vdots \\ y_2(11) \end{pmatrix} \qquad x = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} \qquad A = \begin{pmatrix} x_2(1)^2 & x_2(1) & 1 \\ \vdots & \vdots & \vdots \\ x_2(11)^2 & x_2(11) & 1 \end{pmatrix}$

PCA

The covariance matrices for each of the four matrices D_i is the same:

$$\text{cov}(D_i) = \begin{pmatrix} 11 & 5.5 \\ 5.5 & 4.12 \end{pmatrix}$$

and each has the SVD

$$\text{cov}(D_i) = \begin{pmatrix} 0.87 & -0.48 \\ 0.48 & 0.87 \end{pmatrix} \begin{pmatrix} 14.0 & 0 \\ 0 & 1.08 \end{pmatrix} \begin{pmatrix} 0.87 & 0.48 \\ -0.48 & 0.87 \end{pmatrix}$$

(a) What is the largest eigenvalue of $\text{cov}(D_i)$?
What is the first principal component of D_i?

largest eigen value = 14.0

first principal component = $\begin{pmatrix} 0.87 \\ 0.48 \end{pmatrix}$ ✓

(b) Suppose that we changed the vector $\boldsymbol{x}_1$ by subtracting its mean from each of its entries. How would the first principal component of D_1 change, if at all?

The first principal component of D_1 will not change since all the X_i's are shifted by the same number. The total effect will be no change. ✓

(c) Suppose that we changed the vectors $\boldsymbol{x}_1$ and $\boldsymbol{y}_1$ by multiplying each of their entries by 3. How would the first principal component of D_1 change, if at all?

The first principal component will be multiplied by 3^2

no – only the eigenvalues are scaled.

−1

$\begin{pmatrix} 0.87 \cdot 3^2 \\ 0.48 \cdot 3^2 \end{pmatrix}$ 2 × 1 vector

(d) Suppose that we switched $\boldsymbol{y}_1$ and $\boldsymbol{x}_1$, so that the columns of D_1 are reordered. How would the first principal component of D_1 change, if at all?

The first principal component will not change since the covariance will stay the same.

flipped $\begin{pmatrix} 0.48 \\ 0.87 \end{pmatrix}$

−3

(e) The fourth plot (for D_4) shows a single outlier at the eighth point $(19, 12.5)$. If this (x, y)-pair were removed from D_4, how would the principal components for resulting 10×2 dataset D_4 change?

The principal components will change to other vectors which will have more spread in the data. If the outlier is taken out, there will be much less ~~spread~~ on these axis vectors. If there is more spread in other vectors, the principal components will change accordingly

−2