Biometry 726
Fall 2010
Homework 1
Due Tuesday 14th September 2010

1. 2.3

2. 2.5

3. Read pages 60 - 62 of your text pertaining to the spectral decomposition of a symmetric matrix. Then complete 2.8.

4. 2.11 (Use their hint or try mathematical induction)

5. 2.17

6. For matrix M, we want to find, for each row, which row of matrix M is nearest as measured by Euclidean distance. Recall Euclidean distance between $(x_1 \ x_2 \ \ldots \ x_n)'$ and $(y_1 \ y_2 \ \ldots \ y_n)'$ is defined by

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2}.$$

Write a function that performs this operation and use `apply` to process all rows in the matrix. You may use the `dist` function in `R` or you can write your own distance algorithm. You may also assume there is a unique answer.

7. The figure below shows standardized (i.e. centered to have zero mean and unit standard deviation) values of five variables from the prostate cancer data (posted on the class web site), by normal and high PSA (normal PSA is <10 and high PSA is 10 or more).

(a) Write `R` code to reproduce this graphic. You will want to vertically jitter the points to enhance visibility. Use `lty="b"` to "produce" lines connecting points.

(b) Is there any empirical evidence of association between PSA level and any of the variables based on the plot? Justify your answer (using complete sentences, please).