Problem Set 05

Linear Models – Fall 2015

Due date: 2015-11-11

Problems sets are due at the start of class on the due date. Please hand write or type up and print the solutions; we will not accept e-mail solution sets except in exceptional circumstances. You may discuss problem sets with others, but must write up your own solutions. This means that you should have no need to look at other's final written solutions. Many of these problems come from a variety of textbooks, which are referenced in the problems. These are for citation purposes and not because you will need to consult the text itself (though you may feel free to do so).

I. Singular Value Decomposition

1. The pseudoinverse of a matrix A, denoted A^+, is a generalization of a matrix inverse, which is defined for any matrix A. It has the following four properties:

1. $AA^+A = A$
2. $A^+AA^+ = A^+$
3. $(AA^+)^t = AA^+$
4. $(A^+A)^t = A^+A$

For the singular value decomposition $U\Sigma V^t$ of A, show that $V\Sigma^+U^t$ satisfies all the properties of pseudoinverse, where Σ^+ is the transpose of the rectangular diagonal matrix Σ with all of the non-zero terms replaced with their inverse.

2. An additional way to solve the ordinary least squares problem is to take the pseudoinverse of the design matrix X, and let $\widehat{\beta}$ be equal to X^+y. (a) Explain why this seems reasonable. (b) Show, directly, that the predicted values $X\widehat{\beta}$ will be equal to UU^ty. (c) Apply this technique to solve the ill-conditioned system we observed in class:

```
> X  <- matrix(c(10^9, -1, -1, 10^(-5)), 2, 2)
> beta <- c(1,1)
> y <- X %*% beta
```

(d) Comment on how well this works and why it works well/poorly.

3. The condition number of a matrix can, alternatively, be defined as $||A||_2||A^+||_2$.[1] Verify that this matches the definition given in class as the ratio of the extremal singular values.

[1]The matrix norm $||A||_2$ is the induced norm defined as the supremum of $||Ax||_2/||x||_2$ over all non-zero x.

II. Shrinkage Estimators

1. Calculate the (a) bias, (b) variance, and (c) mean squared error for the ridge regression vector in terms of X, σ^2, and λ. Make sure to justify all of the steps in your derivation.

2. Show that there exists a λ_0 such that the mean squared error of the ridge regression vector $\widehat{\beta}_\lambda$ is lower than that of the ordinary least squares regression vector $\widehat{\beta}$ for all $0 < \lambda < \lambda_0$.

3. Consider again the estimator $\widehat{\beta}_\alpha$ equal to $\alpha \cdot \widehat{\beta}$. Calculate its (a) bias, (b) variance, and (c) mean squared error. Write these in terms of the singular value decomposition of X. How do they compare to the results in question 1?

III. Applications to Image Data

1. Download the Columbia images dataset:

```
http://euler.stat.yale.edu/~tba3/class_data/columbiaImages.zip
```

and metadata

```
http://euler.stat.yale.edu/~tba3/class_data/photoMetaData.csv
```

Read the image `DSC_1739.jpg` into R using the **jpeg** package. Construct a greyscale matrix by averaging the three color channels together and take the singular value decomposition of the matrix. (a) Plot a histogram of the log of the singular values. (b) Now create a new matrix Σ where all but the largest 60 singular values have been set to 0 and reconstruct the matrix with the new Σ. Save the image and open it as a jpeg. (c) How well has it been reconstructed?

2. (a) Now create images with only the top 50, 40, 30, 20, and 10 singular values. (b) How well do these capture the image? (c) What are the last components to disappear? (d) What's the smallest number of singular values that you recommend for producing a thumbnail of the image?

3. Using the image `DSC_1734.jpg`, (a) produce a pixel-by-pixel scatter plot of the red and green channels, and the green and blue channels. Now compute the principal components of this dataset and save three (black and white) images with these components (hint: you will need to scale the values so they fall between 0 and 1). (c) Comment on what is captured in each component.