Homework 1, STAT365

Girish Sastry

January 25, 2011

1 Exercise 2.3

Let m be the median distance between the closest data point and the origin of the unit ball. Thus,

$$Pr[AllNpointsdistance \geq m) = \frac{1}{2}$$

Let x_i be a single point and d_i be its distance to the origin. Then,

$$Pr[d_i \geq m] = 1 - Pr[d_i < r) = 1 - \frac{C\pi m^p}{C\pi 1^p} = 1 - m^p$$

Thus,

$$Pr[AllNpointsdistance \geq m] = \frac{1}{2} = \prod_{i=1}^{N} Pr[d_i > m] = (1 - m^p)^N.$$

Solving for m, we get

$$m = \left(1 - \frac{1}{2}^{\frac{1}{N}}\right)^{\frac{1}{p}}$$

2 Exercise 2.8

In this exercise we compare classification of zipcode data using linear regression and k-nearest neighbors. To calculate the error for linear regression, we sum the difference of squares between $y_{\vec{test}}$ and $\vec{\hat{y}}$ for the test data and the difference of squares between $y_{\vec{train}}$ and $\vec{\hat{y}}$ for the training data, and divide each by the length of the vector to obtain the error. Note that this is equivalent to using R's inbuilt "mean" function. We proceed by a similar error calculation for the k-nearest neighbors algorithm.

Results:

Method	Test Error	Training Error
Linear Regression	3.85%	0.58%
1-NN	2.47%	0.00%
3-NN	3.02%	0.43%
5-NN	3.02%	0.58%
7-NN	3.02%	0.58%
15-NN	3.85%	0.93%

With these results, we see that linear regression performs a little worse than k-NN for small values of k. Both the 1-NN and 3-NN methods give better errors than linear regression. The 5-NN and 7-NN methods have better test error results, and the 15-NN gives the worst results. The large number of predictors may help explain these results. And, as discussed in lecture and in the text, we would expect lower values of k to perform better. However, overall, the model performed well.

3 Exercise 3.3

3.1 Part a

Please refer to Part b, which this justification will be based on. With $a'\beta$, we have a linear combination of β. And, just as in part b, in this case, the sum of the variances and covariances in the linear combination will give us the total variance. And since each v_i and covariance in the linear combination is less than its counterparts in the linear combination for $\hat{\theta}$ (as the variance-covariance matrix exceeds the other variance-covariance matrix by a semidefinite matrix), then the $var(a'\beta) \leq var(a'\hat{\theta})$. Thus, the proof for part b will hold for part a.

3.2 Part b

As in the text, let $\hat{\theta} = CY$ be a different linear estimator of $\hat{\beta}$ and let $C = (X'X)^{-1}X' + D$ where D is a k x n nonzero matrix. Then,

$$\begin{aligned}
V(\hat{\theta}) &= V(CY) = CV(Y)C' = \sigma^2 CC' && (1)\\
&= \sigma^2((X'X)^{-1} + D)(X(X'X)^{-1} + D') && (2)\\
&= \sigma^2((X'X)^{-1}X'X(X'X)^{-1} + (X'X)^{-1}X'D' + DX(X'X)^{-1} + DD') && (3)\\
&= \sigma^2(X'X)^{-1} + \sigma^2(X'X)^{-1}(DX)' + \sigma^2 DX(X'X)^{-1} + \sigma^2 DD' && (4)\\
&= \sigma^2(X'X)^{-1} + \sigma^2 DD' && (5)\\
&= V(\hat{\beta}) + \sigma^2 DD' && (6)
\end{aligned}$$

Because DD' is positive semidefinite,

$$V(\hat{\theta}) \succeq V(\hat{\beta})$$

4 Appendix

This section contains the R code for Exercise 2.8.