Weekly Quiz 7

The **due date** for this quiz is **Sun 10 Mar 2013 3:30 PM GMT**.

Question 1

When the span parameter increases in a loess fit which of the following is true (pick one)?

- ◉ The fit gets more smooth and the bias increases.

- ○ The fit gets more smooth and the variance increases.

- ○ The fit gets more bumpy and the bias increases.

- ○ The fit gets more smooth and fewer points are averaged at each value.

Question 2

Define a data set according to the code

```
set.seed(53535)
xValues = seq(0,2*pi,length=100)
yValues = rnorm(100) + sin(xValues)
```

Fit linear models with the yValues as outcome and a natural cubic spline model for the xValues as the covariates. Fit the model with degrees of freedom equal to each integer between 1 and 10. For each model, calculate the root mean squared error (RMSE) between the fitted values and the observed yValues (the rmse() function in R may help). At what number of degrees of freedom is there the most dramatic drop in the RMSE? Why does this make sense?

- ○ The RMSE drops between df=1 and df=2. This is because the relationship between x and y is not linear.

○

The RMSE drops between df=3 and df=4. This is because the sinusoidal model has one inflection points - like a cubic function.

○

The RMSE drops between df=1 and df=2. This is because there are 2 degrees of freedom in the sinusoidal model.

◉

The RMSE drops between df=2 and df=3. This is because the sinusoidal model has one inflection points - like a cubic function.

Question 3

Load the simpleboot package (you may have to install it first) with the following commands:

```
library(simpleboot)
data(airquality)
attach(airquality)
```

Calculate the 75th percentile of the Wind variable. Then set the seed to 883833 and use the one.boot function with 1,000 replications to calculate the bootstrap standard error of the 75th percentile of the Wind variable.

○ The 75th percentile is: 11.5 The bootstrap s.d. is: 0.5125929

○ The 75th percentile is: 9.7 The bootstrap s.d. is: 0.5965868

◉ The 75th percentile is: 11.5 The bootstrap s.d. is: 0.5965868

○ The 75th percentile is: 11.5 The bootstrap s.d. is: 0.3463986

Question 4

Load the Cars93 data:

```
data(Cars93,package="MASS")
```

Set the seed to 7363 and calculate three trees using the tree() function on bootstrapped samples (samples with replacement). Each tree should treat

the DriveTrain variable as the outcome and Price and Type as covariates.

Predict the value of the following data frame

```
newdata = data.frame(Type = "Large",Price = 20)
```

with each tree and report the majority vote winner along with the percentage of votes among the three trees for that value.

- ⦿ Front Percent of Votes = 100%

- ○ Front Percent of Votes = 66%

- ○ Front Percent of Votes = 80%

- ○ FWD Percent of Votes = 66%

Question 5

Load the vowel.train and vowel.test data sets:

```
library(ElemStatLearn)
data(vowel.train)
data(vowel.test)
```

Set the variable y to be a factor variable in both the training and test set. Then set the seed to 33833. Fit (1) a random forest predictor relating the factor variable y to the remaining variables and (2) an svm predictor using the svm() function in the e1071 package. What are the error rates for the two approaches on the test data set? What is the error rate when the two methods agree on a prediction?

- ⦿ Test error random forest = 0.4199134 Test error svm = 0.3874459 Test error both agree = 0.2823129

- ○ Test error random forest = 2.123354 Test error svm = 2.038695 Test error both agree = 2.123354

- ○ Test error random forest = 0.4199134 Test error svm = 0.3874459 Test error both agree = 0.5714286

○
Test error random forest = 2.123354 Test error svm = 2.038695 Test error both agree = 1.995975

☑ In accordance with the Honor Code, I certify that my answers here are my own work.

Submit Answers   Save Answers