Feedback — Weekly Quiz 6

You submitted this quiz on **Fri 1 Mar 2013 2:37 PM PST**. You got a score of **10.00** out of **10.00**.

Question 1

Which of the following (pick one) is not a step in building a prediction model?

Your Answer		Score	Explanation
⊙ Estimating test set accuracy with training-set accuracy.	✔	2.00	
Total		2.00 / 2.00	

Question 2

If K is small in a K-fold cross validation is the bias in the estimate of out-of-sample (test set) accuracy smaller or bigger? If K is small is the variance in the estimate of out-of-sample (test set) accuracy smaller or bigger. Is K large or small in leave one out cross validation?

Your Answer		Score	Explanation
⊙ The bias is larger and the variance is smaller. Under leave one out cross validation K is equal to the sample size.	✔	2.00	
Total		2.00 / 2.00	

Question 3

Load the South Africa Heart Disease Data and create training and test sets with the following code:

```
library(ElemStatLearn)
data(SAheart)
set.seed(8484)
train = sample(1:dim(SAheart)[1],size=dim(SAheart)[1]/2,replace=F)
trainSA = SAheart[train,]
testSA = SAheart[-train,]
```

Then fit a logistic regression model with Coronary Heart Disease (chd) as the outcome and age at onset, current alcohol consumption, obesity levels, cumulative tabacco, type-A behavior, and low density lipoprotein cholesterol as predictors. Calculate the misclassification rate for your model using this function and a prediction on the "response" scale:

```
missClass = function(values,prediction){sum(((prediction > 0.5)*1) != v
alues)/length(values)}
```

What is the misclassification rate on the training set? What is the misclassification rate on the test set?

Your Answer		Score	Explanation
⊙ Training set misclassification: 0.2727 Test set misclassification: 0.3117	✔	2.00	
Total		2.00 / 2.00	

Question 4

Load the olive oil data using the commands:

```
library(pgmm)
data(olive)
olive = olive[,-1]
```

These data contain information on 572 different Italian olive oils from multiple regions in Italy. Fit a classification tree where Area is the outcome variable. Then predict the value of area for the following data frame using the tree command with all defaults

```
newdata = as.data.frame(t(colMeans(olive)))
```

What is the resulting prediction? Is the resulting prediction strange? Why or why not?

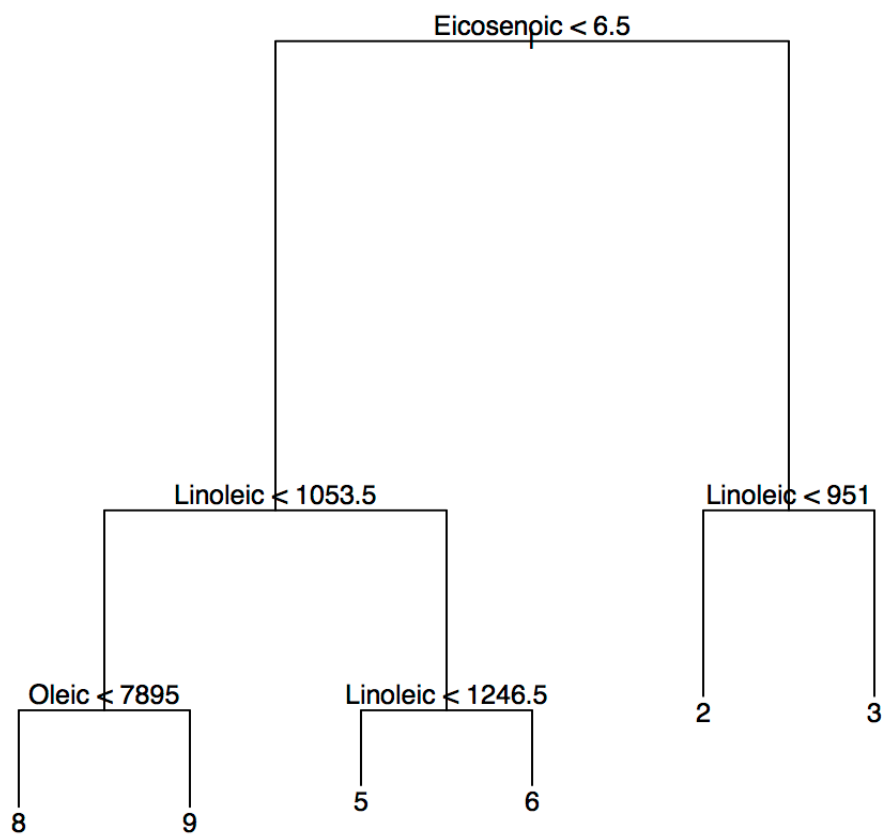| Your Answer		Score	Explanation
⊙ 2.875. It is strange because Region should be a qualitative variable - but tree is reporting the average value of Region as a numeric variable in the leaf predicted for newdata.	✔	2.00	
Total		2.00 / 2.00	

Question 5

Load the olive oil data using the commands:

```
library(pgmm)
data(olive)
olive = olive[,-1]
```

Suppose that I fit and prune a tree to get the following diagram. What area would I predict for a new value of:

```
newData = data.frame(Palmitic = 1200, Palmitoleic = 120, Stearic=200,Ol
eic=7000,Linoleic = 900, Linolenic = 32, Arachidic=60,Eicosenoic=6)
```


Your Answer	Score	Explanation
◉ Area 8	✔ 2.00	

| Total | 2.00 / 2.00 |