Homework 2 Partial Solutions

Problem 1:   For this problem, I was basically looking to see whether everyone just did a proper cross-validation to find the error rates, and that people weren't mixed up between apparent and true error rates. Comparions for 5 classifers are shown below. For $k - nn, k = 1$ gave the best results.

	Accuracy
RANDOM FORESTS	0.99
LDA	0.99
QDA	0.99
KNN	0.77
DECISION TREES	0.88

Problem 2: For this problem, we follow a similar procedure using 10-fold cross validation (although the sample size is large enough to avoid such a method). Comparions for 5 classifers are shown below. For $k - nn, k = 1$ gave the best results.

Random Forests give the best results for Region. Qda seems to be the best for Area. At this point we could simplt use random forests for predicting regions and Qda for predicting Area, but we adopt a different approach. For our predictions, we use the most common one amongst all classifiers. There is one inconsistency between area/region, which is observation six.

	Accuracy
RANDOM FORESTS	1.00
LDA	0.99
QDA	1.00
KNN	0.97
DECISION TREES	0.99

Table 1: Accuracy for Region

	Accuracy
RANDOM FORESTS	0.95
LDA	0.94
QDA	0.97
KNN	0.92
DECISION TREES	0.90

Table 2: Accuracty for Area

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
RANDOM FORESTS	3.00	1.00	2.00	1.00	3.00	1.00	1.00	1.00	3.00	1.00	2.00	3.00	3.00	3.00	1.00	1.00	1.00	3.00	1.00	3.00
LDA	3.00	1.00	2.00	1.00	3.00	3.00	1.00	3.00	3.00	1.00	2.00	3.00	2.00	3.00	1.00	1.00	3.00	3.00	1.00	3.00
QDA	1.00	1.00	2.00	1.00	3.00	1.00	1.00	1.00	1.00	1.00	2.00	2.00	3.00	2.00	1.00	1.00	1.00	3.00	1.00	3.00
KNN	3.00	1.00	2.00	1.00	3.00	3.00	1.00	1.00	3.00	1.00	2.00	3.00	3.00	3.00	1.00	1.00	3.00	3.00	1.00	3.00
DECISION TREES	3.00	1.00	3.00	1.00	3.00	1.00	1.00	1.00	3.00	1.00	3.00	3.00	3.00	3.00	1.00	1.00	1.00	3.00	1.00	3.00
MODE	3.00	1.00	2.00	1.00	3.00	1.00	1.00	1.00	3.00	1.00	2.00	3.00	3.00	3.00	1.00	1.00	1.00	3.00	1.00	3.00

Table 3: Predictions for region

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
RANDOM FORESTS	8.00	3.00	5.00	4.00	7.00	9.00	4.00	1.00	9.00	3.00	6.00	9.00	8.00	9.00	3.00	3.00	3.00	8.00	3.00	7.00
LDA	8.00	2.00	5.00	3.00	7.00	9.00	2.00	9.00	7.00	3.00	9.00	9.00	8.00	7.00	2.00	2.00	8.00	8.00	3.00	7.00
QDA	8.00	4.00	5.00	3.00	7.00	2.00	1.00	3.00	8.00	3.00	5.00	8.00	8.00	8.00	3.00	4.00	5.00	8.00	3.00	7.00
KNN	8.00	3.00	5.00	4.00	7.00	9.00	3.00	1.00	9.00	3.00	5.00	9.00	8.00	9.00	3.00	3.00	8.00	8.00	3.00	7.00
DECISION TREES	7.00	3.00	7.00	2.00	7.00	1.00	3.00	1.00	9.00	5.00	7.00	9.00	8.00	9.00	3.00	3.00	3.00	8.00	3.00	7.00
MODE	8.00	3.00	5.00	4.00	7.00	9.00	3.00	1.00	9.00	3.00	5.00	9.00	8.00	9.00	3.00	3.00	3.00	8.00	3.00	7.00

Table 4: Predictions for area

3