DSCI 35600 – HW 09 (28 pts) Due: Friday, 4/12 Name Key

Problem 1. (10 pts) A dataset `X_train, y_train` contains 200 samples split over 3 classes. A random forest classifier is trained on this data using the following code:

```
random_forest = RandomForestClassifier(n_estimators=3, max_depth=2, bootstrap='True')
random_forest.fit(X_train, y_train)
```

The three trees created by the classifier are described below. Consider a new sample given by `[-1,2,1]`. Find the probability of this sample being in each of the three classes, as predicted by the random forest classifier. That is, find the output of the following code:

```
                0 1 2
Xnew = [[-1,2,1]]
random_forest.predict_proba(Xnew)
```

Provide your answer as an array with shape `(3,)`.

```
* Size: 200 [65, 59, 76], Axis:1, Cut: 0.21
  * Size: 118 [11, 57, 50], Axis:1, Cut: -0.11
    * Size: 99 [10, 53, 36], Predicted Class: 1
    * Size: 19 [1, 4, 14], Predicted Class: 2
  * Size: 82 [54, 2, 26], Axis:0, Cut: 0.06
    * Size: 38 [15, 0, 23], Predicted Class: 2
    * Size: 44 [39, 2, 3], Predicted Class: 0

* Size: 200 [71, 61, 68], Axis:2, Cut: -1.4
  * Size: 48 [12, 25, 11], Axis:0, Cut: -0.02
    * Size: 18 [6, 2, 10], Predicted Class: 2
    * Size: 30 [6, 23, 1], Predicted Class: 1
  * Size: 152 [59, 36, 57], Axis:2, Cut: 2.24
    * Size: 142 [59, 29, 54], Predicted Class: 0
    * Size: 10 [0, 7, 3], Predicted Class: 1

* Size: 200 [62, 75, 63], Axis:0, Cut: -0.94
  * Size: 52 [8, 25, 19], Axis:0, Cut: -2.66
    * Size: 5 [1, 0, 4], Predicted Class: 2
    * Size: 47 [7, 25, 15], Predicted Class: 1
  * Size: 148 [54, 50, 44], Axis:0, Cut: 2.06
    * Size: 124 [43, 47, 34], Predicted Class: 1
    * Size: 24 [11, 3, 10], Predicted Class: 0
```

$p_1 = [0.3947, 0, 0.6053]$

$p_2 = [0.4155, 0.2042, 0.3803]$

$p_3 = [0.1489, 0.5319, 0.3191]$

$p = [0.3197, 0.2454, 0.4349]$

Problem 2. (6 pts) Assume we are creating a model for performing multi-class classification on a dataset for which there are 9 categorical labels.

a) If a one-versus-one (OVO) classification scheme is used, how many binary classification models must be created?

$\binom{9}{2} = \frac{9(8)}{2} = 9(4) = \boxed{36}$ (one for each pair of classes)

b) If a one-versus-rest (OVR) classification scheme is used, how many binary classification models must be created?

$\boxed{9}$ (one for each class.)

Problem 3. (6 pts) Five probability models are used to create a single ensemble model for a classification task with 4 categorical labels. A single sample is fed into the ensemble, generating the following probability distributions:

	Class 0	Class 1	Class 2	Class 3
Model 1	0.1	0.5	0.1	0.3
Model 2	0.2	0.1	0.3	0.4
Model 3	0.1	0.4	0.2	0.3
Model 4	0.1	0.45	0.1	~~0.36~~ 0.35
Model 5	0.3	0.15	0.1	0.45

a) Assuming a soft-voting scheme is used, find the probability distribution that this model would return for this sample.

[0.16, 0.32, 0.16, 0.36]

b) Which class would the ensemble predict for this sample if soft-voting is used?

Class 3

c) Which class would the ensemble predict for this sample if hard-voting is used?

Class 1

Problem 4. (6 pts) Write **True** or **False** next to each of the following statements.

This problem will be graded as follows: +1 points for each correct answer, -1 points for each incorrect answer, and 0 points for each blank answer. So, for instance, one correct answer and one incorrect answer will cancel each other out. You cannot get less than zero points on this problem, even if every answer is incorrect.

- When using bagging or pasting to create an ensemble model, each model in the ensemble is trained on the entire training set. F
- When using bagging, sampling from the training set is performed with replacement. T
- When using the RandomForestClassifier class to create an ensemble of decision trees, each tree in the ensemble is trained on a different subset of the training set. T
- Assume we are using the RandomForestClassifier class to create an ensemble of decision trees. When constructing a particular tree model in the ensemble, each splitting step is performed on the feature that will generate the best results at that step. F
- It is important to perform feature scaling when creating an SVM model with the radial basic function as the kernel. T
- Hard-voting can only be used in an ensemble model in which each of the individual models is capable of producing probability estimates. F