

## EE2211 Tutorial 10

### Question 1:

We have two classifiers showing the same accuracy with the same cross-validation. The more complex model (such as a 9<sup>th</sup>-order polynomial model) is preferred over the simpler one (such as a 2<sup>nd</sup>-order polynomial model).

- a) True
- b) False

### Question 2:

We have 3 parameter candidates for a classification model, and we would like to choose the optimal one for deployment. As such, we run 5-fold cross-validation.

Once we have completed the 5-fold cross-validation, in total, we have trained \_\_\_\_\_ classifiers. Note that, we treat models with different parameters as different classifiers.

- A) 10
- B) 20
- C) 25
- D) 15

### Question 3:

Suppose the binary classification problem, which you are dealing with, has highly imbalanced classes. The majority class has 99 hundred samples and the minority class has 1 hundred samples. Which of the following metric(s) would you choose for assessing the classification performance? (Select all relevant metric(s) to get full credit)

- a) Classification Accuracy
- b) Cost sensitive accuracy
- c) Precision and recall
- d) None of these

### Question 4:

Given below is a scenario for Training error rate  $Tr$ , and Validation error rate  $Va$  for a machine learning algorithm. You want to choose a hyperparameter ( $P$ ) based on  $Tr$  and  $Va$ .

P	Tr	Va
10	0.10	0.25
9	0.30	0.35
8	0.22	0.15
7	0.15	0.25
6	0.18	0.15

Which value of P will you choose based on the above table?

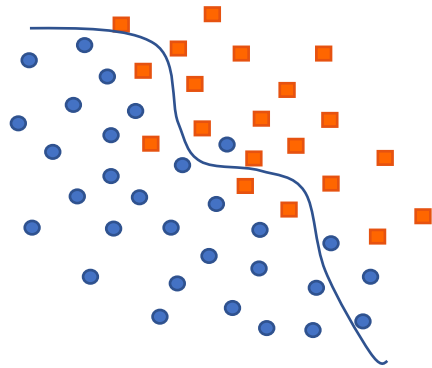
- a) 10
- b) 9
- c) 8
- d) 7
- e) 6

(Binary and Multicategory Confusion Matrices)

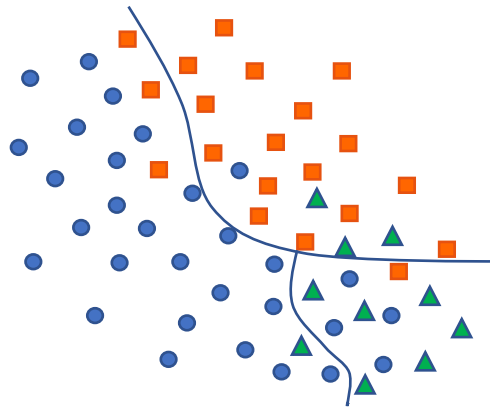
**Question 5:**

Tabulate the confusion matrices for the following classification problems.

(a) Binary problem (the class-1 and class-2 data points are respectively indicated by squares and circles)



(b) Three-category problem (the class-1, class-2 and class-3 data points are respectively indicated by squares, circles and triangles).



(5-fold Cross-validation)

**Question 6:**

Get the data set “from sklearn.datasets import load\_iris”. Perform a 5-fold Cross-validation to observe the best polynomial order (among orders 1 to 10 and without regularization) for validation prediction. Note that, you will have to partition the whole dataset for training/validation/test parts, where the size of validation set is the same as that of test. Provide a plot of the average 5-fold training and validation error rates over the polynomial orders. The randomly partitioned data sets of the 5-fold shall be maintained for reuse in evaluation of future algorithms.