CS 4824 / ECE 4424 Homework 3 (Written Portion) Solutions Total: 25 points

Question 1 (Deep Learning True/False) [8 points]

a. True,
b. False (if we use ReLU at the output layer, the outputs will no longer be bounded within 0 to 1),
c. True (the derivative of ReLU is discontinuous at 0),
d. False (the primay motivation is to avoid the vanishing gradient problem).

Question 2. (Support Vector Machines True/False) [4 points]

a. False (SVM will choose the one that maximizes the margin).
b. True

Question 3. (Ensemble Learning Matching) (4 points)

a. iii; b. i., c. ii., d. iv.

Question 4 (Class Imbalance) [9 points]

a)

[0.5 points] Precision = 90/(90 + 10) = 0.90,
[0.5 points] Recall = TPR = 90/(10 + 90) = 0.90,

[0.5 points] FPR = 10/(10 + 90) = 0.10.

b)

[0.5 points] Precision = 90/(90 + 90) = 0.50,

[0.5 points] recall = TPR = 90/(10 + 90) = 0.90,

[0.5 points] FPR = 90/(90 + 810) = 0.1

c)

[0.5 points] Recall, TPR and FPR, have remained the same.

[0.5 points] Precision has changed.

[1 point] Even though the classifier was still performing the same as measured by TPR and FPR, precision is sensitive to the relative number of members of each class.

d)

[1 point] Specificity = 0.90 for both tables,

[1 point] Sensitivity = 0.90 for both tables

e)

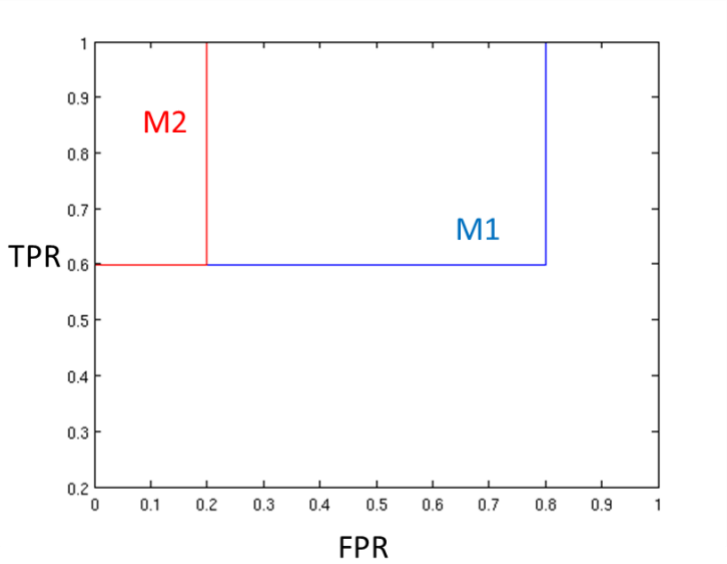
[1 point] (Sensitivity, specificity) measures are preferred in the medical domain because unlike (precision, recall) they are not sensitive to the relative number of subjects in each class. Thus, if a drug is evaluated at two different sites, but with different relative numbers in each class, the results will still be comparable.

[1 point] (Precision, recall) measures are preferable when the number of false alarms must be minimized, such as in fraud detection.

Question 5 (ROC Curve) [Not graded]

a)

ROC Curve for M2 (red) and M1(blue) is shown below



The area under the curve (AUC) of M1 = 0.68 and the AUC of M2 = 0.92. Since M2 has a higher AUC than M1, M2 is a better model. Also notice that for any given FPR, M2 has better (or same) TPR than M.

b)

For M1 at t = 0.4, Precision = 0.5, Recall = 0.6, F-Measure = 0.54.

For M2 at t = 0.4, Precision = 1, Recall = 0.6, F-Measure = 0.75.

By looking at the F-Measure values, we conclude that M2 is a better model, since it has a higher F-Measure than M1. This result is consistent with the result in (a), by using AUC of the ROC curve as the evaluation criterion.

c)

For M1 at t = 0., Precision = 1, Recall = 0.6, F-Measure = 0.75.

For M2 at t = 0.7, Precision = 1, Recall = 0.2, F-Measure = 0.33.

By looking at the F-Measure values, we conclude that M1 is a better model, since it has a higher F-Measure than M2. This result is inconsistent with the result in (a), by using the AUC of the ROC curve as the evaluation criterion. This indicates that evaluating a classifier by computing its F-Measure at a given threshold can give a different result than by constructing its ROC curve and measuring its AUC.

Practice Questions

Question 6. (Support Vector Machines)
Yes M_B would be exactly same as M_A and support vectors of M_B would also be exactly same as that of M_A. This is because the learning of SVM only depends on the support vectors, which either lie on the margin hyperplanes or have non-zero slack variables. Removing the non-support vectors from the training set has no effect on SVM learning as all of the information necessary to discriminate the classes is captured in the support vectors.

Question 7. (Ensemble Learning True/False)

a. False, b. True, c. False

Question 8. (Ensemble Learning)

Similarity: Both are ensemble learning methods.
Difference: The ensemble classifiers in bagging are independent to each other, whereas in boosting, the instances that are incorrectly predicted by previous classifiers are given more weight. Also, bagging only tries to reduce the variance of ensemble classifiers, while boosting tries to reduce the bias as well as the variance.

Question 9. (Class Imbalance)

a)

The expected value of TPR and FPR for classifier C1 is p. The expected value of TPR and FPR for classifier C2 is 2p. Both classifiers are just randomly assigning test points to the positive class irrespective of their true class. This corresponds to the random guessing line in an ROC curve with TPR = FPR.

b)

The precision of both C1 and C2 is 100/(100 + 900) = 1/10. The recall of C1 and C2 are p and 2p respectively. Hence, if we consider precision and recall, C2 appears to show better {precision, recall} values than C1, although both classifiers are basically randomly guessing. Hence, in this particular case, {TPR, FPR} provides a better indication of the relative performance of classifiers C1 and C2?

Question 10. (Class Imbalance)

a)

Accuracy = 0.87

Precision = 0.2

Recall = 0.22

F-measure = 0.21

b)

Accuracy is a poor indicator of overall performance of the algorithm. In the case of imbalanced class problem, accuracy does not evaluate well the performance of the algorithm on the smaller class.

F-measure is a good indicator of classifier's performance, since it takes into account both precision and recall.

c)

Classify all as negative class:

Accuracy = 0.93