CS 4824 / ECE 4424 **Homework 3 (Written Portion)**

Note: Please prepare your solutions electronically using Word/Latex/any other software of our choice. Hand-written submissions are not always legible and thus difficult to grade.

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

State whether the following statements are true or false. No explanation needed.

a. Sigmoid activation function, when used at the hidden layers, makes deep neural networks prone to vanishing gradient problem.
b. Saturation of output problem can be avoided by using the ReLU activation function at the output layer.
c. The ReLU activation function is non-differentiable at 0.
d. The primary motivation behind using the ReLU activation function is to avoid problem of overfitting in deep learning.

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

State whether the following statements are true or false. No explanation needed.

a. For the linearly separable case, SVM and perceptron will always result in the same linear decision boundary.
b. Hard-margin SVM will fail to provide any solution if the training set is not linearly separable, i.e., there does not exist any linear hyperplane that discriminates all positive points from all negative points.

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

On the left, you are given four classification scenarios that are possible for different settings of bias and variance of the classifier. On the right, you are given four types of classifiers. You need to match the classification scenarios to their corresponding classifiers.

Classification Scenarios	**Classifier Types**
(a) Low Bias, Low Variance	(i) Underfitting Classifier
(b) High Bias, Low Variance	(ii) Overfitting Classifier
(c) Low Bias, High Variance	(iii) Ideal Classifier
(d) High Bias, High Variance	(iv) Worst Classifier

Question 4 (Class Imbalance) [14 points]

a) **[2 points]** You are working with a doctor to evaluate how well a new, inexpensive blood test can detect a particular type of cancer. 1000 subjects are recruited from a population at high risk for the cancer and evaluated for cancer using a very expensive, but 100% accurate medical procedure. 100 subjects are found to have cancer. The 100 subjects with cancer and another 100 subjects without cancer are given the inexpensive blood test. Results are shown in the following confusion matrix.

	Predicted by Blood Test		
Actual	Cancer	No Cancer	
Cancer	90	10	100
No Cancer	10	90	100
	100	100	200

Using the above confusion matrix, compute the precision, recall, TPR, and FPR for the Cancer class.

b) **[2 points]** The doctor is very excited about these results but wants to see what the results will be after all the blood tests are evaluated. The remaining 800 subjects (none of which have cancer) are given the blood test.

The confusion matrix for all 1000 subjects is given below.

	Predicted by Blood Test		
Actual	Cancer	No Cancer	
Cancer	90	10	100
No Cancer	90	810	900
	180	820	1000

For this new confusion matrix, compute the precision, recall, TPR, and FPR for the Cancer class.

c) **[3 points]** Which of the measures (precision, recall, TPR, FPR) have changed and which have stayed the same? Comment on why some measures were affected and others were not.

d) **[3 points]** You were disappointed by the change you observed in one of the measures from the first confusion matrix to the second, but the doctor was not. The doctor tells you to consider sensitivity (Recall) and specificity (1-FPR). Compute the sensitivity and specificity for both confusion matrices. Do their values change from the first confusion matrix to the second?

e) **[4 points]** When would we prefer (Sensitivity, specificity) as the preferred choice of evaluation measures? When would we prefer (precision, recall) as the preferred choice of evaluation measures?

Question 5 (ROC Curve) [16 points]

You are asked to evaluate the performance of two classification models, M1 and M2, for a binary classification problem with classes ‘+’ and ‘–’. For every test instance, x, each of the two models provides a posterior probability of x belonging to class ‘+’. The following table provides a list of 10 test instances with their true classes, and their posterior probabilities of belonging to class ‘+’, according to M1 and M2.

Instance	True Class	P(+|M1)	P(+|M2)
1	+	0.98	0.27
2	+	0.31	0.45
3	+	0.92	0.95
4	+	0.31	0.46
5	+	0.93	0.23
6	-	0.33	0.13
7	-	0.47	0.08
8	-	0.46	0.19
9	-	0.24	0.37
10	-	0.45	0.04

a) **[5 + 3 points]** Plot the ROC curve for both M1 and M2 (you should plot them on the same graph.) (*Hint:* to obtain the ROC curve of a model, you need to compute the TPR and FPR for every threshold on the posterior probability estimated by the model.) Using the ROC curves, compute the AUC for M1 and M2. Which model do you think is better? Explain your reasons.

b) **[4 points]** Suppose you choose a cutoff threshold to be t = 0.4 for both the models, M1 and M2. In other words, any test instance whose posterior probability is greater than t will be classified as a positive example. Compute the Precision, Recall, and F-Measure for M1 and M2 after using the cutoff threshold of t = 0.4. Which model is better using F-measure as the evaluation criterion? Are the results consistent with what you expect from the ROC curve?

c) **[4 points]** Repeat part (b) using t = 0.7. Which model is better using F-measure as the evaluation criterion? Are the results consistent with what you expect from the ROC curve?

Practice Questions

Question 6. (Support Vector Machines)
You are given a training set A comprising of 200 data points with binary class labels. You use this data set to train an SVM model, M_A. After training, you find that only 10 data points are support vectors for M_A, while the remaining 190 data points are all non-support vectors. You remove 100 such non-support vectors from data set A to produce a smaller data set, B (of size 100). If you train an SVM model, M_B, on this new data set B, would you expect M_A to be equal to M_B? Also, would M_B have the same set of support vectors as M_A? Briefly explain.

Question 7. (Ensemble Learning True/False)

For a binary classification problem, you are given a collection of base classifiers where every base classifier has an error rate of ϵ. The ensemble prediction is simply the majority vote of the predictions of the base classifiers. State whether the following statements are “true” or “false.” No explanation needed.

a) If $\epsilon = 0.5$ and the predictions of all base classifiers are independent, the error rate of the ensemble classifier will be smaller than ϵ.
b) If $\epsilon = 0.3$ and the predictions of all base classifiers are independent, the error rate of the ensemble classifier will be smaller than ϵ.
c) If $\epsilon = 0.3$ and all base classifiers are identical, the error rate of the ensemble classifier will be smaller than ϵ.

Question 8. (Ensemble Learning)

State one similarity and one difference between bagging and boosting.

Question 9. (Class Imbalance)

Consider a test data of 1000 samples with two classes: +ve class (100 samples) and –ve class (900 samples). We have two random classifiers C1 and C2. Classifier C1 classifies test data randomly to +ve class with a probability p (and to –ve class with probability 1-p) and classifier C2 classifies test data to +ve class randomly with a probability 2p.

a) What is the expected value of TPR and FPR for classifiers C1 and C2? Is C2 a better classifier than C1?

b) What is the expected value of precision and recall for classifiers C1 and C2? Which evaluation metric pair between {TPR and FPR} and {precision and recall} do you think is correctly indicating the relative performance of classifiers C1 and C2?

Question 10. (Class Imbalance)

You are given a classification algorithm that predicts whether it will rain tomorrow (+) or not (-). The confusion matrix of the algorithm on 1000 test days is given below:

		Predicted	
Actual		+	-
	+	20	50
	-	80	850

a) Compute accuracy, precision, recall, and F-measure with respect to ‘+’ class.

b) Which of these metrics is a poor indicator of the overall performance of your algorithm? Which of these metrics is a good indicator of the overall performance? Give a one sentence reason why this is the case?

c) Construct a trivial classifier that achieves better accuracy by classifying all test instances to the same class, irrespective of their attributes.