Note: All coding problems to be submited with Github Link. Do not Upload the files/folder. Use git commands only.

Note: this is the distribution of questions:

(a) Question 1 to Question6: Required for everyone.

(b) Question 7 - Question 8: Required only for Graduate Students

Problem 1 (5 points)

Accuracy is probably the most commonly used metric in classification problems. Which statement below is True?:

(a) Accuracy is very easy to interpret and is aligned with what you want to measure.

(b) Accuracy is non-differentiable and cannot be used for direct optimization via gradient descent.

(c) We want accuracy to count every error as equal.

(d) If the number of samples in each category is comparable, we cannot make any mistake on the less populated category.

Problem 2 (10 points)

You are working as a Machine Learning Engineer in Metflix Inc. You are building a model to classify users who watch a lot of movies in Ultiverse. What metrics will you choose to evaluate your model?

Problem 3 (10 points)

Which method is used involved in numerical optimization of an appropriate selection of model criterion? How do you define the error of such estimator?

Problem 4 (5 points)

We covered Automatic Differentiation in class. Consider the following function:

$$f(x, y, z) = \frac{1}{3}(x_1 x_2 \sin(x_3) + exp^{x_1 x_2}) \tag{1}$$

(a) Draw a computation graph for this function.

(b) provide the computation trace of the function

(c) provide the forward AD trace for $x_1 = 1$, $x_2 = 0$, $x_3 = 0$

(d) provide the reverse mode AD trace for $x = 1$, $y = 0$, $z = 0$

List the detailed computation steps (the trace) for forward and backward mode of AD. Provide your answers the same way we did in class, by using notations like:

$$v_{-2}, v_{-1}, v_0, v_1, v_2 \tag{2}$$

to get you started:

$$v_{-2} = 1, \dot{v}_{-2} = \frac{\partial v_{-2}}{\partial x_{-2}} = 0 \tag{3}$$

Provide all the steps like this and values for other nodes in computation graph.

Problem 5 (10 points)

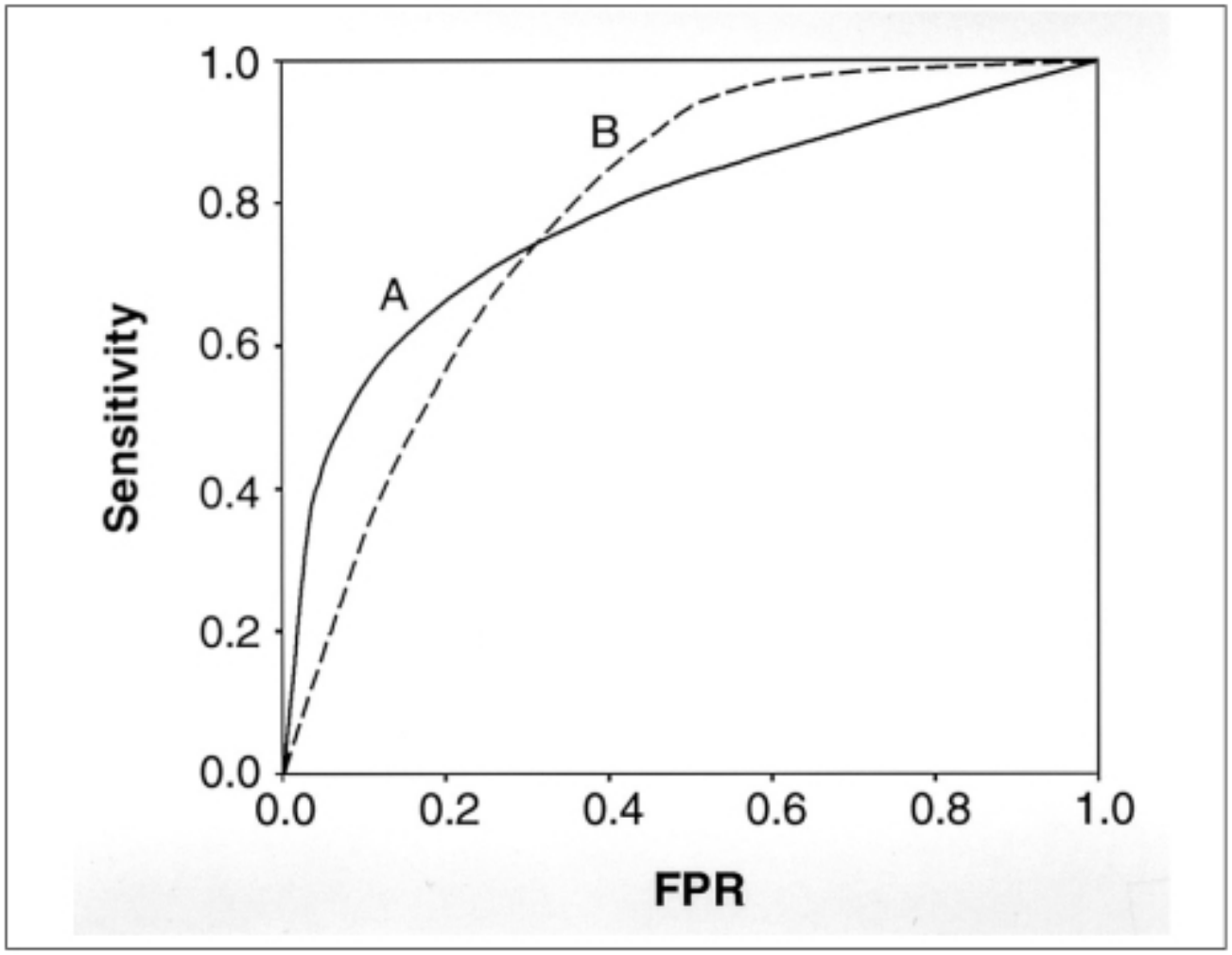

Figure 1: Two curves with equal area under the ROC curve

In the above figure we have ROC curves for two classifiers (A and B) which have equal areas under the curve (AUC).

(a) Which classifier is better among these two? (write your detailed thinking)

(b) Describe a situation in which undoubtedly classifier A should be preferred.

(c) Describe a situation in which undoubtedly classifier B should be preferred.

(d) Which factors determine the area under the curve?

Problem 6 (20 points)

Technically we need to compute the gradient with respect to $\mathbf{W}_i$, the linear transform (or parameter) matrix (tensor) for layer i. Yet, we are computing $\frac{\partial \ell}{\partial \boldsymbol{x}}$, gradient of the loss ℓ with respect to the input $\boldsymbol{x}$. How come?

——————————— Bonus for undergraduates beyond this line

Problem 7 (20 points)

Compare the following metrics and explain which one is better.

(a) ROC and AUC

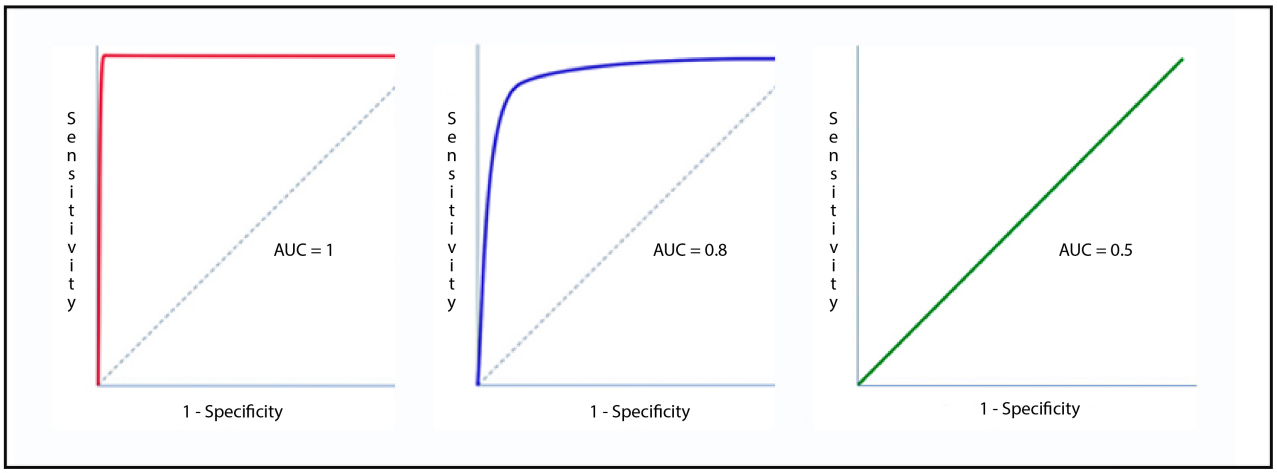

Figure 2: Accuracy

(b) For the same figure, explain relevance to True Positive Rate (TPR) and False Positive Rate (FPR)
