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

(c) Question 9- Question 10: Bonus marks

Problem 1 (5 points)

We studies 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 the computation graph for this equation.

(b) List the detailed computation steps 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 2 (5 points)

Select the correct option.Accuracy is probably the most commonly used metric in the most common type of Deep learning problems:

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

(b) Accuracy is differential and can be used to directly in gradient based optimisation process

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

(d) If the number of sample is similar, we cannot make any mistake on the less populated category.

Problem 3 (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 4 (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 5 (10 points)

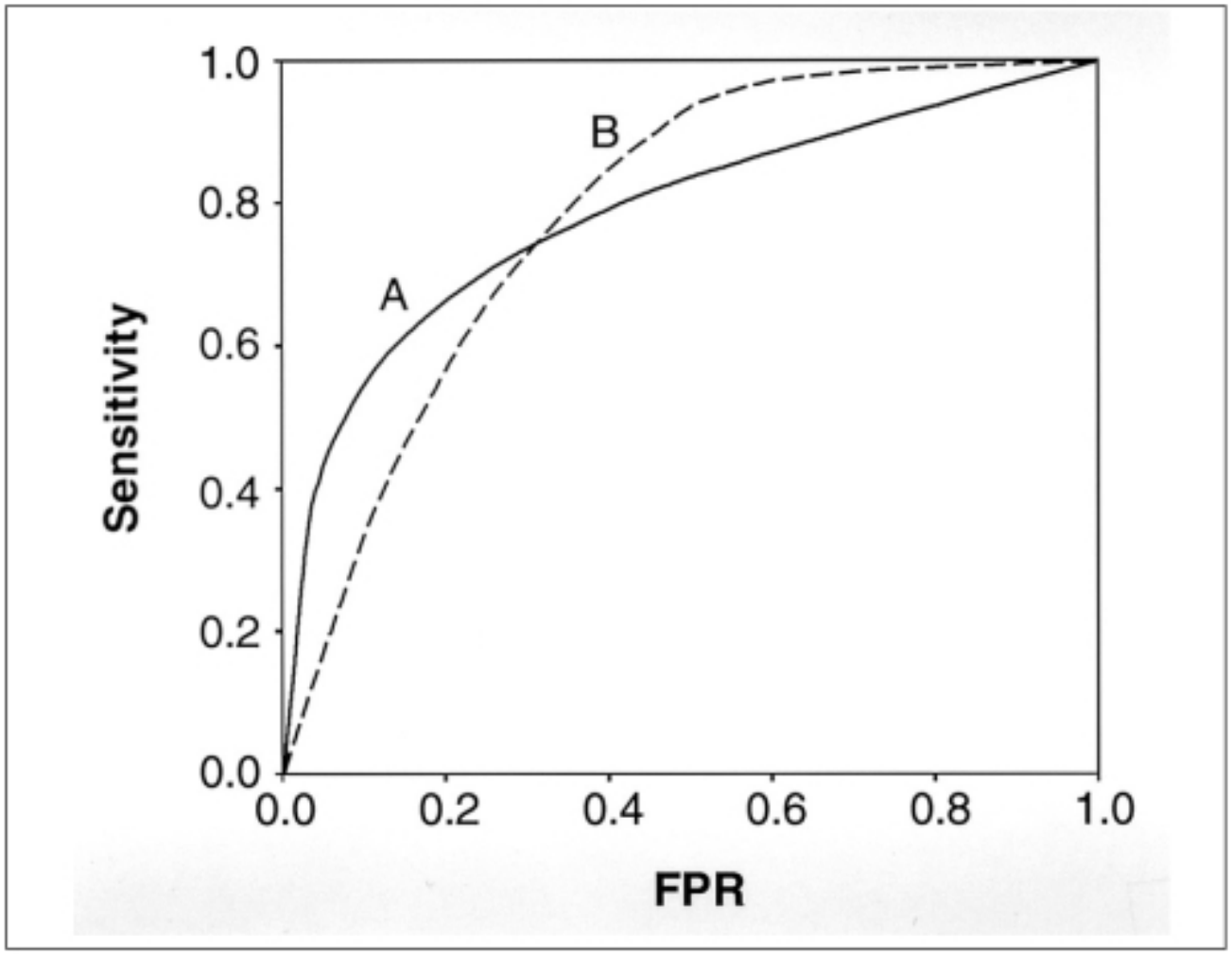

Figure 1: Two curves with equal area under the ROC curve

In the above figure we have two ROC curve (A and B) which has the same area under the curve. Which one is better among these two and what governs the area under the curve?

Problem 6 (20 points)

why is it that technically we're looking for gradient with respect to W_i, where W_i is the linear transform (or parameter) matrix (tensor) for each layer. Yet, we are computing $\frac{d(loss)}{d(x)}$, where x is the input. 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

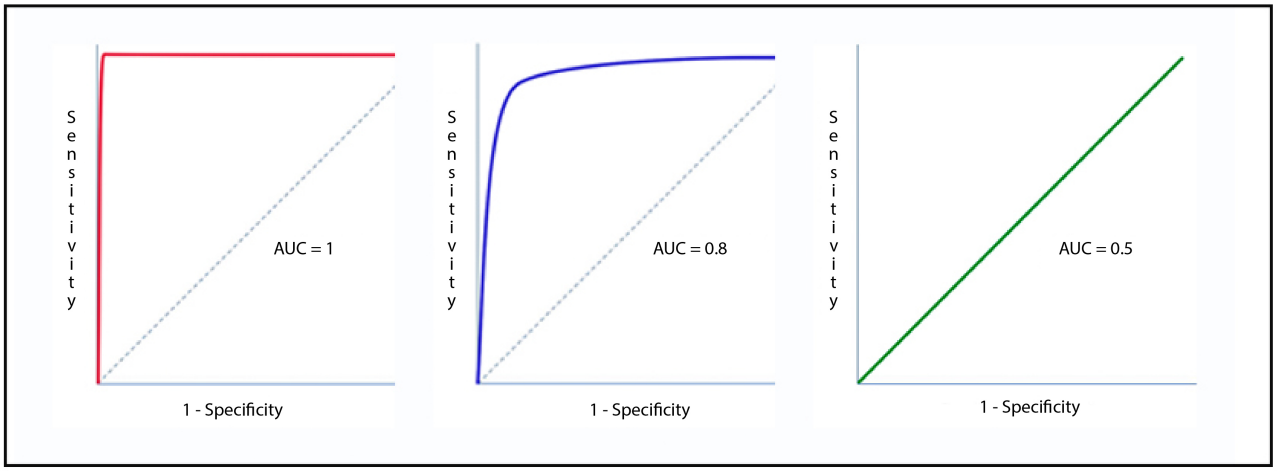

Figure 2: Accuracy

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

Problem 8 (extra credit 20 points)

Learn a simple and efficient procedure for obtaining analytic derivatives of complex matrix/vector objectives. Based on the exposition in Appendix III of https://arxiv.org/abs/1911.04048 manually differentiate the following objective function:

$$\frac{1}{4n(n-1)}\mathbf{1}^T \log\left((\boldsymbol{x}\odot\boldsymbol{x})\mathbf{1}^T + \mathbf{1}(\boldsymbol{x}\odot\boldsymbol{x})^T - 2\boldsymbol{x}\boldsymbol{x}^T + \mathbf{I}\right)\mathbf{1},$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{1}$ is appropriately sized vector of ones, and $\odot$ denotes Hadamard element-wise product. List each and every one of your step. You may type them down by hand on paper very clearly and submit a photo. **The answer expression for $\dot{\boldsymbol{x}}$ must be typeset electronically though**