CS5590, CS3390, AI5000, AI2000: Quiz-1

6-Sep-2022, Tuesday, 7pm-8:30pm

ROLL NO.______________________

Important Instructions

1. PLEASE WRITE YOUR ROLL NO correctly IN THE BLANK ABOVE.
2. In the questions below, fill the blanks/boxes in this booklet itself such that the respective statements become **true**.
3. While filling the blanks/boxes strictly follow the formatting **instructions** in the respective question (if any).
4. In case a box is given for filling a detailed answer, then marks will be awarded **if and only if** the justification is precise.
5. Begin attempting the problems in **rough sheets** first. Then, fair copy your answers into this question paper while respecting the boundaries of the blanks/boxes. Hand-writing must be **neat** and **legible**.
6. The space provided in blank/box is more than enough with legible font size. If you are feeling the space is too less, then it simply means there is a more concise way of writing the answer. The evaluator shall be **strictly** following the policy of ignoring writings outside the blanks/boxes and those that are too tiny to read!
7. After the exam, please submit this question paper booklet to the invigilator. Do not submit your rough work.
8. There will be **no** partial marking. Be very cautious to write the entire mathematically expression correctly. It is not enough to write 95% of expression/formula correctly. You may loose all marks for the blank/box even if you forget a '-' or '1/2' etc.

Questions

1. Consider the logistic regression setting taught in the lecture. Here, the learning task is a ______________________________ problem [[*Fill this blank with either "regression" or "multi-class classification" or "binary classification". **1/4 mark***]]. Let $\phi : \mathcal{X} \mapsto \mathbb{R}^n$ be a given feature map. The model employed in logistic regression is the ________________ [[*Fill this blank with the appropriate proper noun. **1/4 mark***]]. The mathematical definition of this model is given by the expression:

   $$\rule{8cm}{0.4pt}.$$

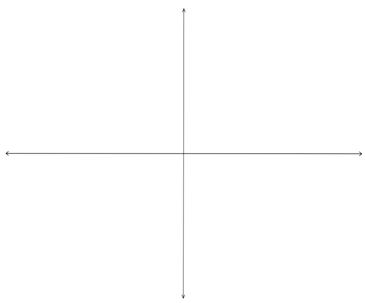
   [[*In the above blank, use 'w' to denote the parameter of this model. **1/2 mark***]]

   The mathematical expression for the loss function used here is:

   $$l(w, x, y) \equiv \rule{3.5cm}{0.4pt}.$$

   [1 mark]

   This loss function can be visualized using the plot below:

   [[*Fill the plot appropriately to roughly depict the graph of logistic loss. Clearly label the axes, without which no marks will be given. **1/2 mark***]].

   If p^* is the underlying (unknown) likelihood relating the inputs and labels, then, the Bayes optimal, restricted to the functions in the model, is given by the mathematical expression:

   $$f^* = \rule{5.5cm}{0.4pt}$$

[[*No marks will be given if you write general expressions for Bayes optimal. You need to write the specific expression for the logistic regression set-up.* ***1 mark***]]. Let the training data be $\mathcal{D} \equiv \{(x_1, y_1), \ldots, (x_m, y_m)\}$. The name of the important assumption that relates p^* to $\mathcal{D}$ is ____________________. [[*Fill this blank with the appropriate proper noun.* ***1/2 mark***]]. This assumption formally means the following:

[**1/2 mark**]

The ERM problem in this case is the following mathematical optimization problem:

____

[**1/2 mark**]

If the ERM solution is denoted by $\hat{w}_m$, then the label for any $x \in \mathcal{X}$ shall be computed using the formula: ____________.

[**1/2 mark**]

Now, say, the training data actually is

$$\mathcal{D} = \left\{ \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, 1 \right), \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, -1 \right), \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}, -1 \right), \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, 1 \right) \right\}$$

and the feature map ϕ is defined by

$$\phi\left(\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \right) \equiv ((z_1 + z_2)\%2) - \frac{1}{2} \ \forall \ z_1, z_2 \in \mathbb{R}.$$

Here, $a\%b$ is the remainder when a is divided by b. For this case, in the box below, write down the optimal solution[1] of the ERM problem along with justification:

[1] You are welcome to solve this optimization problem in any way you prefer. For e.g., analytically, manually iterating through gradient descent etc.

[1 mark]

For this specific training data and feature map, suppose we wish to perform linear classification using the 0-1 loss. Then, run the perceptron algorithm in rough using manual calculations and write all parameter iterates until convergence including initialization in the box below. For each iterate write down the update equation too. No other details are required

[1 mark]

2. Consider the linear regression setting taught in lectures with training data as: $\mathcal{D} = \{(2,1),(4,5)\}$ (usual convention of set of input,label pairs). Consider the feature map $\phi(x) = x$. Analytically solve the ERM problem in rough work and write down the final ERM solution in this blank: $w_{\phi}^{ERM} =$ __ [[***1/2 mark***]]. With this solution, the explained variance computed on the training set is _____ [[*Fill the blank with appropriate number.* ***1/2 mark***]]. Now, consider another feature map, $\psi(x) \equiv \begin{bmatrix} x \\ 1 \end{bmatrix}$. With this feature map, analytically solve the ERM problem in rough work and write down the final ERM solution in this blank: $w_{\psi}^{ERM} =$ _____ [[***1 mark***]]. With this solution, the explained variance computed on the training set is __ [[*Fill the blank with appropriate number.* ***1/4 mark***]]. Now, consider another feature map, $\nu(x) \equiv \begin{bmatrix} x \\ 1 \\ -1 \end{bmatrix}$. With this feature map, analytically solve the ERM problem in rough work and write down the final ERM solution you obtained in this blank: $w_{\nu}^{ERM} =$ _____

[[*1 mark*]]. With this solution, the explained variance computed on the training set is __ [[*Fill the blank with appropriate number. 1/4 mark*]].

3. In the lectures you were taught how to model the Bayes optimal in a binary classification task using linear functions (over input feature space). Now suppose you have a multi-class classification problem with 3 classes: 'Ξ', '‡', and '¶'. However, still you are only allowed to use the linear model taught in lectures. Think about how you can model the Bayes optimal in a 3-class classification task using these linear functions. With this way of modelling the Bayes optimal in mind, according to you, the loss function, l, appropriate for this task would be defined by $l(w, x, \Xi) \equiv$ ________________, $l(w, x, \ddagger) \equiv$ ________________, $l(w, x, \P) \equiv$ ________________. Here, w denotes the parameter of the linear model.

[**1.5 Marks**]

Observe that your way of modelling the Bayes optimal with linear functions has an inherent ('wrong'?) bias. More specifically, if the parameter changes a little then the label for a fixed x changes preferentially to one of the other two classes. In this sense, there is am implicit (unequal) nearness between different class pairs.

Now, suppose you are allowed to model functions of the form $f(x) = W^\top \phi(x)$, where W is $n \times 3$, where ϕ is a feature map. You may use the notation $W = [w_1\ w_2\ w_3]$, where $w_i \in \mathbb{R}^n$. Think about how you can model the Bayes optimal in a 3-class classification task using these '3-dimensional linear functions'. With this way of modelling the Bayes optimal in mind, according to you, the loss function, l, appropriate for this task would be defined by $l(W, x, \Xi) \equiv$ ________________, $l(W, x, \ddagger) \equiv$ ________________ $l(W, x, \P) \equiv$ ________________.

[**2.5 Marks**]