CS5590: Exam - 1

7:45pm-9:15pm, 03-Sep-2019

ROLL NO._____

Note: Fill in the blanks/boxes appropriately such that the respective statements become true. While filling the blanks/boxes strictly follow the instructions in the respective question appearing immediately after/before the blank/box. You are free to use any standard mathematical symbols like $\pi, e, \Sigma, \|\cdot\|, \log, \max$ etc. Answers that are not simplified enough, (correct) answers in wrong format, illegible writings, and those outside the blanks/boxes, will be ignored by the evaluator. Please attempt the problems in rough sheets first and prepare answers for all the blanks/boxes in rough. Then fair copy them in this sheet while respecting the boundaries of the blanks/boxes.

1. Consider a binary classification problem with input space $\mathcal{X} = \mathbb{R}^n$, and output space $\mathcal{Y} = \{-1, 1\}$. Consider a training set given by $= \{(x_0, 1), (-x_0, -1)\}$, where $x_0 \in \mathcal{X} (\neq 0)$ is a given fixed point. It is proposed to employ the logistic loss and the linear inductive bias (without the norm-bound). Then, the simplified expression for the ERM problem is:

$$\min_{w \in \mathbb{R}^n} \left[\right].$$

Your expression in the previous blank must involve w, x_0 only.

[1 Mark]

In the box below argue that this optimization problem as no solution:

[1 Mark]

Now, the inductive bias is changed to linear functions with norm-bound given by hyperparameter $W = 1$. Then, the optimal solution, $\hat{w}$, of the corresponding ERM problem in it's original form involving the hyperparameter[1] W is given by: $\hat{w} = $ _____. This expression must involve x_0 alone.

[0.5 Mark]

The above exercise highlights yet another advantage of the norm-bounded linear functions over the set of all linear functions!

---

[1]Here, the ERM is not re-written in the Tikhonov form i.e., it is NOT rewritten in the regularized risk minimization form.

2. Assume you have a classification problem with 3 classes: '✠', '†', and '❡'. The loss function, l, you would employ if you were restricted to model only one real-valued function, f, is given by $l(x, ✠, f) \equiv$ _____ , $l(x, †, f) \equiv$ _____ , $l(x, ❡, f) \equiv$ _____ .

[1.5 Marks; Practice set problem!]

Suppose you re allowed to model 3 real-valued functions, say, f, g, h. Then, the loss function you would employ is given by: $l(x, ✠, (f, g, h)) \equiv$ _____ .

[2 Marks; Practice set problem!]

3. Consider a regression problem where the input space, $\mathcal{X} = \mathbb{R}^n$, and the output space, $\mathcal{Y} = \mathbb{R}$. It is proposed to use the inductive bias as the set of all affine functions:

$$\mathcal{G} \equiv \left\{ g \mid \exists\, w \in \mathbb{R}^n,\ b \in \mathbb{R}\ \ni\ g(x) = w^\top x - b\ \forall\ x \in \mathcal{X} \right\}.$$

The parameters are $w \in \mathbb{R}^n, b \in \mathbb{R}$. The loss to be used is the square loss. Let the input vectors in the training set be arranged as column vectors in the matrix $X_{n \times m}$, where m is the training set size. Let $y_{m \times 1}$ denote the vector with entries as the corresponding outputs in the training set. Let us denote the q-dimensional vector with all entries as unity by 1_q. Then, the simplified expression for the ERM optimization problem written in terms of $X, y, 1_m, w, b$ is given by:

$$\min_{w \in \mathbb{R}^n, b \in \mathbb{R}} \left[\frac{1}{m} \left(\underline{\hspace{6cm}} \right) \right].$$

Your expression in the previous blank must not use explicit symbols for columns, entries of X, y. In other words, please employ vector operations rather than scalar ones.

[1 Mark]

Now, if the inductive bias is changed to norm-bounded affine functions:

$$\mathcal{G}_W \equiv \left\{ g \mid \exists\, w \in \mathbb{R}^n,\ b \in \mathbb{R},\ \|w\| \leq W\ \ni\ g(x) = w^\top x - b\ \forall\ x \in \mathcal{X} \right\},$$

the simplified expression for the ERM problem in Tikhonov form[2] turns out to be:

$$\min_{w \in \mathbb{R}^n, b \in \mathbb{R}} \left[\underline{\hspace{8cm}} \right].$$

Your expression in the previous blank must be in terms of $X, y, 1_m, w, b, \lambda$ only, where λ is the hyperparameter (that replaces W).

[0.5 Mark]

Now, by repeating the analysis done in the lecture for the case of linear/ridge regression, or otherwise, find an analytical expression for the optimal solution, $(\hat{w}, \hat{b})$, of this problem. The optimal $\hat{w}$ satisfies the following linear equalities:

$$\left(\underline{\hspace{8cm}} \right) \hat{w} = \left(\underline{\hspace{6cm}} \right)$$

Your expression for the first of the previous two blanks must be only in terms of $X, \lambda, I_n, 1_m, m$, where I_n is the identity matrix of size n. And, the second must be in terms of $X, y, 1_m, m$.

---

[2] Regularized risk minimization form.

In the following box, please write a formal proof of why the matrix in the first of the previous two blanks, denoted by, say, P, is positive definite.

| | |
| | |

Hence, the optimal $\hat{w} = P^{-1}q$, where q denotes the vector in the second of the previous two blanks.

[1 Mark]

The optimal $\hat{b}$ is given by the expression: _____ . This expression must involve $\hat{w}, X, y, 1_m, m$ alone.

[0.5 Mark]