IT 585 - Advanced Machine Learning

End Semester Exam

May 05, 2023
Duration: 150 minutes
Maximum Marks: 40

Note: This is an open notebook exam. You can have your notebook/handwritten notes/printed copies of handwritten notes with you. However printed scribe notes/textbooks are not allowed. In case you have any doubt, make an appropriate assumption, state the assumption clearly, and proceed. Proofs should be complete. Answer to the point. Verbose answers without relevant content will be penalized.

1. Consider the perceptron in two dimensions: $h(\mathbf{x}) = sign(\mathbf{w}^{\mathbf{T}}\mathbf{x})$ where $\mathbf{w} = [w_0, w_1, w_2]^T$ and $\mathbf{x} = [1, x_1, x_2]^T$. Technically, $\mathbf{x}$ has 3 coordinates, but we call this perceptron two-dimensional because the first coordinate is fixed at 1. Show that the regions on the plane where $h(\mathbf{x}) = 1$ and $h(\mathbf{x}) = -1$, are separated by a line. (Just an explanation is OK). If we express this line by the equation $x_2 = ax_1 + b$,what are the slope a and the intercept b in terms of w_0, w_1 and w_2 **[5 Marks]**

2. Changing the error measure can change the result of the learning process. Suppose you have N data points $x_1, x_2, \ldots, x_N$ and you wish to estimate a 'representative' value. Prove that if your algorithm is to find the hypothesis h that minimizes the in sample sum of squared deviations,

$$E_{\text{in}}(h) = \sum_{n=1}^{N}(h - x_n)^2,$$

   then your estimate will be the in-sample mean. **[3 Marks]**

3. Let $\mathcal{X}$ be the Boolean hypercube $\{0,1\}^n$. For a set $I \subseteq \{1,2,...,n\}$ we define a parity function h_I as follows. On a binary vector $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \{0,1\}^n$

$$h_I(\mathbf{x}) = (\sum_{i \in I}(x_i)) \bmod 2$$

   That is, h_I computes parity of bits in I. Find the VC-dimension of the class of all such parity functions, $\mathcal{H}_n$-parity$= \{h_I : I \subset \{1,2,\ldots,n\}\}$. [***Hints***: *i)Consider the total no. of possible points and ii) Show a set of size equal to the VC-dimension which can be shattered*] **[4+4 Marks]**

4. Suppose that we do a coin tossing experiment 5 times and get the following output (T,T,H,H,T). The output of each toss is a Bernoulli random variable with unknown parameter θ which represents the probability of a heads.

   (a) Find the MLE for θ i if θ is unrestricted i.e lies in $(0, 1)$.

   (b) Suppose we impose the restriction that $\theta \in \{0.2, 0.5, 0.7\}$. Find the MLE for θ in this case.

   (c) Assume θ is restricted as in part. But now we have discrete priors for different values of θ given as $\pi_\theta(0.2) = 0.1$, $\pi_\theta(0.5) = 0.01$, and $\pi_\theta = 0.89$. Find the MAP estimate for θ.

[2+3+3 Marks]

5. (a) Consider the kernel $K(\mathbf{x}, \mathbf{z}) = \mathbf{x}^T\mathbf{z} + 4(\mathbf{x}^T\mathbf{z})^2$ where the vectors $\mathbf{x}$ and $\mathbf{z}$ are 2-dimensional. This kernel is equal to an inner product $\phi(\mathbf{x})^T\phi(\mathbf{z})$ for some definition of ϕ. What is the feature mapping ϕ? **[4 Marks]**

   (b) Assume that $n = 4$, and that $\mathbf{x}_1 = [1, 1]^T, \mathbf{x}_2 = [2, 2]^T, \mathbf{x}_3 = [-1.5, -1.5]^T$ and $\mathbf{x}_4 = [4, 4]^T$. We now train an SVM classifier with bias w_0, and in addition with slack variables. Show that for any labeling of the four training examples, the optimal parameter vector $\mathbf{w} = [w_1, w_2]^T$ has the property that $w_1 = w_2$. [***Hint:*** *Think about the optimal* $\mathbf{w}$ *in terms of the support vector coefficients, datapoints and the target labels.*] **[4 Marks]**

6. Imagine a binary classification problem where training data are given by

$$\mathcal{D} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)\}$$

where, $y_i \in \{+1, -1\}$, and $\mathbf{x}_i \in \mathbb{R}^d$. Let us define

$$\begin{aligned}
n_+ &= \text{Number of positive examples} \\
n_- &= \text{Number of negative examples} \\
\mathbf{m}_+ &= \frac{1}{n_+} \sum_{i|y_i=+1} \mathbf{x}_i \\
\mathbf{m}_- &= \frac{1}{n_-} \sum_{i|y_i=-1} \mathbf{x}_i
\end{aligned}$$

Consider the following classification rule which assigns the class label to any given test example $\mathbf{x}$.

$$f(\mathbf{x}) = \begin{cases} 1 & \text{if } \|\mathbf{x} - \mathbf{m}_+\| < \|\mathbf{x} - \mathbf{m}_-\| \\ 0 & \text{otherwise} \end{cases}$$

Show that this classifier can equivalently be written as follows

$$f(\mathbf{x}) = sign\left((\mathbf{m}_+ - \mathbf{m}_-)^\top \mathbf{x} + \frac{1}{2}\left(\|\mathbf{m}_-\|^2 - \|\mathbf{m}_+\|^2\right) \right)$$

[8 Marks]