Boston University

ENG EC 414 Introduction to Machine Learning

Exam 1

Released on Monday, 5 October, 2020 (120 minutes, 42 points + 2 bonus points), submit to Gradescope

---

- *There are 6 problems plus 1 bonus one.*
- *For each part, you must clearly outline the key steps and provide proper justification for your calculations in order to receive full credit.*
- *You can use any material from the class (slides, discussions, homework solutions, etc.), but you cannot look for solutions on the internet. Also, be aware of the limited time.*

Problem 1.1 [5pts] Let $f(z) := z^2$ and $\mathcal{A} := [-1, 1]$. Compute: $\underset{z \in \mathcal{A}}{\operatorname{argmin}} \frac{1}{13+\sqrt{1+2 \cdot f(z)}}$.

Problem 1.2 [6pts] Let $\mathcal{Y} := \{1.5,\ \ 2.0,\ \ 3.5,\ \ 6.0\}$ and $y_1 = y_2 = 1.5,\ \ y_3 = 2.0,\ \ y_4 = y_5 = 3.5,\ \ y_6 = 6.0$.

(a) [2pts] Compute $\underset{y \in \mathcal{Y}}{\operatorname{argmin}} \frac{1}{6} \sum_{j=1}^{6} \mathbf{1}[y \neq y_j]$.

(b) [2pts] Compute $\underset{y \in \mathbb{R}}{\operatorname{argmin}} \frac{1}{6} \sum_{j=1}^{6} (y - y_j)^2$.

(c) [2pts] Compute $\underset{y \in \mathbb{R}}{\operatorname{argmin}} \frac{1}{6} \sum_{j=1}^{6} |y - y_j|$.

Problem 1.3 [10pts]

(a) [2pts] Consider the hyperplane parametrized by $\mathbf{w}$ and b with $b = 3$ and $\mathbf{w} = (1,\ -4,\ 8)^\top$. Determine which of the following points lie on the hyperplane: (i) $\mathbf{x}_1 = (-2,\ 2,\ 1)^\top$, (ii) $\mathbf{x}_2 = (0,\ 1,\ 0)^\top$, (iii) $\mathbf{x}_3 = (1,\ 3,\ 1)^\top$.

(b) [2pts] Compute the distance of $\mathbf{x}_4 = (-1,\ -1,\ -1)^\top$ from the hyperplane in part (a).

(c) [3pts] Compute the orthogonal projection of the point $\mathbf{x}_4$ from part (b) onto the hyperplane in part (a).

(d) [3pts] Determine parameters $\mathbf{w}$ and b of the hyperplane passing through the following 3 points: $\mathbf{x}_5 = (1/2,\ 0,\ 0)^\top$, $\mathbf{x}_6 = (1,\ 1,\ 0)^\top$, $\mathbf{x}_7 = (-1,\ 1,\ -1)^\top$.

Problem 1.4 [6pts] Consider the following set of feature vectors and corresponding real-valued labels

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} -3 \\ 1 \end{bmatrix}, \mathbf{x}_4 = \begin{bmatrix} 4 \\ 3 \end{bmatrix}, \quad y_1 = 4, y_2 = 2, y_3 = -8, y_4 = 2\ .$$

(a) [4pts] Fix $b = 0$ and compute by hand the ordinary least squares (OLS) solution $\mathbf{w}^\star$.

(b) [2pts] Compute the OLS prediction of $\mathbf{w}^\star$ and $b = 0$ for the vector $\mathbf{x}_1$.

Problem 1.5 [7pts] Let $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)$ be a training set with feature vectors $\mathbf{x}_j \in \mathbb{R}^d$ and labels $y_j \in \mathbb{R}$. Consider the following cost function for Regularized Least Square without bias, that is, there is no b:

$$g(\mathbf{w}) = \|\mathbf{w}\|^2 + \frac{1}{2m} \sum_{j=1}^{m} (y_j - \mathbf{x}_j^\top \mathbf{w})^2 .$$

<u>Note that this formulation is slightly different from the one seen in class, don't just copy from the slides!</u>

(a) [2pts] Compute the gradient $\nabla g(\mathbf{w})$.

(b) [2pt] Provide pseudocode for an algorithm to minimize $g(\mathbf{w})$ based on gradient descent with zero initialization, a fixed positive step size $\eta > 0$, and the maximum number of iterations T.

(c) [3pt] After a certain number of iterations less than the maximum number of iterations, $\boldsymbol{w}_t$ in gradient descent stops changing, that is $\boldsymbol{w}_{t+1} = \boldsymbol{w}_t$. Can it happen? If yes, in which situations? If no, why?

Problem 1.6 [8pts] Consider the following training set of feature vectors and corresponding binary labels

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} -3 \\ 1 \end{bmatrix}, \mathbf{x}_4 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad y_1 = -1, y_2 = 1, y_3 = 1, y_4 = -1 .$$

(a) [2pts] Hand-plot the training set. Proper labeling of axes and key points is needed to receive full credit.

(b) [2pts] Is it possible to find a hyperplane that linearly separates this training set? A motivation for your answer is needed to receive full credit.

(c) [2pts] Will the Perceptron converge on this dataset? A motivation for your answer is needed to receive full credit.

(d) [2pts] Using the usual augmentation to include the bias in features and hyperplane, compute by hand the first update $\tilde{\boldsymbol{w}}_2$ of the Perceptron algorithm starting from $\tilde{\boldsymbol{w}}_1 = [0, 0, 0]^\top$, after seeing the example $\tilde{\mathbf{x}}_1 = \begin{bmatrix} 1 \\ \mathbf{x}_1 \end{bmatrix}$.

Problem 1.7 [Bonus, 2pts] Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ equal to $f(x_1, x_2) = \frac{1}{2}x_1^2 + \frac{1}{5}x_2^2 + \frac{1}{4}\sin(2x_1)$. Is it convex? Motivate your answer.