Math 123 **Problem Set 6** **Due : Thursday 04/28 at 11:59 p.m. EST**

Instruction: Read the homework policy. You should submit a PDF copy of the homework and any associated codes on Gradescope. Your PDF must be a single file and not multiple images.

1. We study the least squares problem for the case where all observations might not be equally reliable. The standard least square formulation assumes that each data point is equally reliable as the other. In certain applications, due to the nature of acquisition of data, this assumption is violated. In this setting, the observations can be assigned different weights. A larger weight indicates that a data point in consideration is more reliable. To integrate the weight in the least square formulation, define the following inner product: $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{W}^T \mathbf{W} \mathbf{y}$ where $\mathbf{W}$ is a nonsingular matrix of weights, $\mathbf{x}$ is the vector of least squares coefficients, and $\mathbf{y}$ is the dependent variable/response vector.

(a) Following the derivation of the normal equations for the least squares problem, that we did in class, show that the weighted least square solution is equivalent to solving the weighted normal equation: $\mathbf{A}^T \boldsymbol{W}^T \boldsymbol{W} \mathbf{A} \mathbf{x} = \mathbf{A}^T \boldsymbol{W}^T \boldsymbol{W} \mathbf{b}$?

(b) Find the weighted least-squares solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \qquad \mathbf{W} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Plot the data and the least-squares fit on a graph.

2. Consider the optimization problem corresponding to the soft SVM problem,

$$\min_{\mathbf{w} \in \mathcal{R}^d, b} \frac{1}{2} ||\mathbf{w}||^2 + C \sum_{i=1}^{n} \xi_i \quad \text{subject to} \quad y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0 \text{ for } i = 1, ..., n$$

We consider the effect of the regularization parameter C.

(a) What kind of problem do we obtain as $C \to \infty$? Justify your answer.

(b) For very small C, what could you say about the margin? Briefly explain your answer.

(c) For very large C, what could you say about the margin? Briefly explain your answer.

3. Given n points and their labels $(\mathbf{x}_i, y_i)$ for $1 \leq i \leq n$ with each $\mathbf{x}_i \in \mathcal{R}^d$, consider the optimization problem corresponding to the hard margin support vector machines classification problem.

$$\min_{\mathbf{b}, \mathbf{w} \in \mathcal{R}^d} \frac{1}{2} \mathbf{w}^T \mathbf{w} \quad \text{subject to} \quad y_j(\mathbf{w}^T \mathbf{x}_j + b) \geq 1 \text{ for } j = 1, ..., n \tag{1}$$

Let $d = 2$. Let $\mathbf{x}_1 = (0, 0)$, $\mathbf{x}_2 = (2, 2)$, $\mathbf{x}_3 = (2, 0)$ and $\mathbf{x}_4 = (3, 0)$ be the training points with labels $y_1 = -1$, $y_2 = -1$, $y_3 = 1$ and $y_4 = 1$ respectively.

(a) Derive the optimal solution to (1) i.e. find the equation of the optimal hyperplane. (*Follow these steps:* Define Lagrangian, calculate partial derivatives of Lagrangian, define the dual problem for SVM, solve dual problem to find Lagrange multipliers, substitute into dual problem to find weights, substitute into a support vector to find bias). Then, report α, $\mathbf{w}$ and b from your code/derivation.

(b) Based your solution in (a), find an explicit form for the classifier function.

(c) Using the classifier in (b), find the labels of the test points $(6, 2)$ and $(1, 1)$.

(d) Compute the geometric margin.

(e) What are the equations of the support vectors?

Remark: Show detailed work. Answers that simply guess the the equation of the optimal hyperplane receive no credit for (a). To get full credit, you need to solve the problem by considering the optimization objective and the constraints of the dual form in (1). If necessary, use built-in constrained optimization functions (for Python "Scipy.optimize"; for Matlab "fmincon") of your choice and report the results in (a). If you have code for (a), please submit it along with your answer.

4. Download the heart attack analysis and prediction dataset `heart.csv` from `https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset`. Pick one feature of your choice and apply multiple linear regression or classification (using logistic regression), as approrpiate, to predict that feature. Assess the validity of your model and interpret your results. [You can use an in-built function, and to choose any appropriate validation measures.]