CSCI 416/516 Homework #2

DUE: October 25, 2023, at 11:59 pm

~~DUE: October 18, 2023, at 11:59 pm~~

CSCI 416/516: Each Problem begins with an allocation of points, represented as [u pts/g pts]. If you are registered in CSCI 416, you can receive up to u pts on this Problem; if you are registered in CSCI 516, you can receive up to g pts on this Problem. The last Problem is optional for undergraduates (CSCI 416) but required for graduates (CSCI 516). **Write down which session you are in / are you a graduate or undergraduate student**.

Submission: You need to submit both your homework report (answers to the Problems) as a pdf file through Blackboard. Please show the work on how you reach the conclusion for each question, if applicable.

- **Problem 1 [3pts/3pts]: Linear Regression.**
  What is the objective function (loss function) in linear regression?

- **Problem 2 [3pts/3pts]: Linear Regression.**
  Given the following data: $x = [1, 2, 3, 4, 5]$, $y = [2, 3, 4, 5, 6]$, calculate the slope (weight) w and y-intercept (bias) b for the best-fit line.

- **Problem 3 [3pts/2pts]: Gradient Descent.**
  The gradient, given the cost function $\mathcal{J}$ and the weight matrix $\boldsymbol{w} = [w_1, ..., w_D]$ where D is the weight matrix's dimensionality, is described as $\nabla_{\boldsymbol{w}}\mathcal{J}$. What is the expanded form of $\nabla_{\boldsymbol{w}}\mathcal{J}$ expressed in terms of partial derivatives regarding $\mathcal{J}$ and $\boldsymbol{w}$?

- **Problem 4 [3pts/2pts]: Logistic Regression.**
  Given the predicted label y and the true label t of a datapoint $\boldsymbol{x}$, what is the formula for cross-entropy loss, $\mathcal{L}_{\mathrm{CE}}$? Why do we replace the square error loss $\mathcal{L}_{\mathrm{SE}}$ for classification (with linear models), favoring the cross-entropy loss instead?

- **Problem 5 [3pts/2pts]: Support Vector Machine.**
  We want to maximize the margin between the cluster of positive samples and the cluster of negative samples using SVM. Suppose the support vectors in the cluster of positive samples fall on the vertical line $\boldsymbol{\theta}^\top \boldsymbol{x}_+ = 1$, and the support vectors in the cluster of negative samples fall on the vertical line $\boldsymbol{\theta}^\top \boldsymbol{x}_- = -1$, what is the formula, expressed in terms of $\boldsymbol{\theta}$, that we want to maximize?

- **Problem 6 [bonus 3pts/3pts]: Support Vector Machine.**
  The optimization objective of SVM is given as $\min_{\boldsymbol{\theta}} C\sum_{i=1}^{N}[y_i \text{cost}_1(\boldsymbol{\theta}^\top \boldsymbol{x}_i)+(1-y_i)\text{cost}_0(\boldsymbol{\theta}^\top \boldsymbol{x}_i)]+\frac{1}{2}\sum_{j=1}^{d}\theta_j^2$, where cost_0 and cost_1 are defined using the hinge loss. Explain the difference between the scenario in which the tunable hyperparameter C is large and the scenario in which C is small - what are we favoring, by making C large or small?