

Introduction to Machine Learning (CSCI-UA.473): Homework 2

Instructor: Lerrel Pinto

September 20, 2022

Submission Instructions

You must typeset the answers using L^AT_EX and compile them into a single PDF file. Name the pdf file as $\langle \text{Your-NetID} \rangle_{\text{hw2.pdf}}$ and the notebook containing the coding portion as $\langle \text{Your-NetID} \rangle_{\text{hw2.ipynb}}$. The PDF file should contain solutions to both the theory portion and the coding portion. Submit the files through the following Google Form - <https://forms.gle/Rf63VnEaMoLcWr7p7>. The due date is **October 4, 2022, 11:59 PM**. You may discuss the questions with each other but each student must provide their own answer to each question.

Questions

Question 1: Empirical vs. Expected Cost (10 points)

We approximate the true cost function with the empirical cost function defined by:

$$\mathbb{E}_x [E(g(x), f(x))] = \frac{1}{N} \sum_{i=1}^N E(g(x^i), y^i), \quad (1)$$

where N is the number of training samples, f is the unknown function, g is the learnable function, E is the cost function, y^i is the label associated with the input x^i . In Eq. 1, the left-hand side of the equation represents the expected value of the cost between $g(x)$ and $f(x)$ for every x in the dataset, and the right-hand side approximates this expectation by computing a mean over the errors assigning equal weight to each sample. In the above equation is it okay to give an equal weight to the cost associated with each training example? Given that we established that not every data x is equally likely, is taking the sum of all per-example costs and dividing by N reasonable? Should we weigh each per-example cost differently, depending on how likely each x is? Justify your answer.

Question 2: Simple Linear Regression Model (10 points)

Consider the following model: $Y_i = 5 + 0.5X_i + \epsilon_i$, $\epsilon_i \stackrel{iid}{\sim} N(0, 1)$

1. What is $\mathbb{E}[Y|X = 0]$, $\mathbb{E}[Y|X = -2]$ and $Var[Y|X]$?
2. What is the probability of $Y > 5$, given $X = 2$?
3. If X has a mean of zero and variance of 10, what are $\mathbb{E}[Y]$ and $Var[Y]$?
4. What is $Cov(X, Y)$?

Question3: Least Squares Regression (10 points)

Consider the linear regression model:

$$y = \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_k x_k + \epsilon, \epsilon \quad (2)$$

where y is a dependent variable, x_i corresponds to independent variables and θ_i corresponds to the parameters to be estimated. While approximating a best-fit regression line, though the line is a pretty good fit for the dataset as a whole, there may be an error between the predicted value \hat{y} and true value y for every data point $\mathbf{x} = [x_1, x_2, \dots, x_k]$ in the dataset. This error is captured by $\epsilon \sim N(0, \sigma^2)$, where for each data point with features x_i , the label \hat{y} is drawn from a Gaussian with mean $\theta^T \mathbf{x}$ and variance σ^2 . Given a set of N observations, provide the closed form solution for an ordinary least squares estimate $\hat{\theta}$ for the model parameters θ .

For the ordinary least squares method, the assumption is that $Var(\epsilon_i|X_i) = \sigma^2$, where σ is a constant value. However, when $Var(\epsilon_i|X_i) = f(X_i) \neq \sigma^2$, the error term for each observation X_i has a weight W_i corresponding to it. This is called Weighted Least Squares Regression. In this scenario, provide a closed form weighted least squares estimate $\hat{\theta}$ for the model parameters θ .

Question 4: Linear vs Logistic Regression (5 points)

Explain. with equations, the difference between linear and logistic regression.