

New York University Tandon School of Engineering
Computer Science and Engineering

CS-UY 4563: Written Homework 1.
Due Thursday, February 6th, 2020, 11:59pm.

Collaboration is allowed on this problem set, but solutions must be written-up individually. Please list the names of any collaborators at the top of your solution set, or write “No Collaborators” if you worked alone.

For just this first problem set, 10% extra credit will be given if solutions are typewritten (using LaTeX, Markdown, or some other mathematical formatting program).

Problem 1: Practice Framing a Supervised Learning Problem (10pts)

A university admissions office wants to predict the success of students based on their application material. They have access to past student records to learn a good algorithm.

- (a) To formulate this as a supervised learning problem, identify a possible target variable. This should be some variable that measures success in a meaningful way and can be easily collected (in an automated manner) by the university. There is no one correct answer to this problem.
- (b) Is the target variable continuous or discrete-valued?
- (c) State one possible variable that can act as the predictor for the target variable you chose in part (a).
- (d) Without looking at data, would a linear model for the data be reasonable? If so, what sign do you expect the slope to be? If not, what might be a better model?

Problem 2: Practice Minimizing a Loss Function (10pts)

Consider a linear model of the form

$$f_{\beta}(x) = \beta x,$$

which is the same as the linear model we saw in class, but with the intercept forced to zero. Such models are used when we want to force the predicted value $f_{\beta}(x) = 0$ when $x = 0$. For example, if we are modeling y = output power of a motor vs. x = the input power, we would expect $x = 0 \Rightarrow y = 0$.

- (a) Given data $(x_1, y_1), \dots, (x_n, y_n)$, write the equation for a loss function which measures prediction accuracy using the sum-of-squared distances between the predicted values and target values.
- (b) Derive an expression for the β that minimizes this loss function. Do you get the same expression that we got for β_1 in the full linear model?

Problem 3: Building Intuition about Different Loss Functions. (10pts)

Suppose we have data $y_1, \dots, y_n \in \mathbb{R}$ and we want to choose some value $m \in \mathbb{R}$ which is “most representative” of our dataset. This is sometimes called the “central tendency” problem in statistics. A machine learning approach to this problem would measure how representative m is of the data using a loss function.

- (a) Consider the loss function $L(m) = \sum_{i=1}^n (y_i - m)^2$. Show that $L(m)$ is minimized by setting $m = \bar{y}$, where $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ is the *mean* of our data.
- (b) Consider the loss function $L(m) = \max_i (y_i - m)$. What value of m minimizes this loss? Hint: Using derivatives will not help here – try just thinking about the minimization problem directly.
- (c) **Bonus (5pts):** Consider the loss function $L(m) = \sum_{i=1}^n |y_i - m|$. Show that $L(m)$ is minimized by setting m to the *median* of the data.
- (d) In a few short sentences, discuss when you might prefer each of the three losses above. Is the median typically considered a more “robust” measure of central tendency than the mean? Why?

Problem 4: Practice with Non-linear Transformations.

(**Hint:** Take a look at the example at the end of my Lecture 2 notes)

A medical researcher wants to model, $f(t)$, the concentration of some chemical in the blood over time. She believes the concentration should decay exponentially in that

$$f(t) = z_0 e^{-\alpha t}, \tag{1}$$

for some parameters z_0 and α . To confirm this model, and to estimate the parameters z_0, α , she collects a large number of time-stamped samples (t_i, c_i) , $i = 1, \dots, n$, where c_i is the measured concentration at time t_i . Unfortunately, the model (1) is non-linear, so she can't directly apply the linear regression formula to estimate z_0 and α .

- (a) Taking logarithms, show that we can transform our training data so that the conjectured relationship between predictor and target variables is in fact linear.
- (b) Write pseudocode (or actual Python) for how you might estimate z_0 and α using this transformation.