Introduction to Machine Learning (CSCI-UA.473): Homework 1

Instructor: Sumit Chopra

September 14^{th}, 2021

Submission Instructions

You must typeset the answers using LaTeXand compile them into a single PDF file. Name the pdf file as: ⟨Your-NetID⟩_hw1.pdf. For the programming part of the assignment, complete the Jupyter notebook named HW1.ipynb. Create a ZIP file containing both the PDF file and the completed Jupyter notebook. Name it ⟨Your-NetID⟩ hw1.zip. Submit the ZIP file on Brightspace. The due date is **September 27th, 2021, 11:59 PM**.

Theory

Question T1: Empirical vs. Expected Cost (10 points)

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

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

where N is the number of training samples, f is the unknown function, g is the learnable function, y^i is the label associated with the input x^i. 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 T2: Perceptron Learning Algorithm (10 points)

The weight update rule of the Perceptron Learning Algorithm (PLA) is given by:

$$w(t+1) \leftarrow w(t) + y(t)x(t). \tag{2}$$

Prove the following statements:

1. Show that $y(t)w^T(t)x(t) < 0$ (2 points)
2. Show that $y(t)w^T(t+1)x(t) > y(t)w^T(t)x(t)$ (4 points)
3. Argue that the move from $w(t)$ to $w(t+1)$ is the right move as far as classifying $x(t)$ is concerned. (4 points)

Question T3: Gradient of Logistic Regression (10 points)

The logistic regression loss for a single sample (x, y) can be written as

$$\mathcal{L}_w(x, y) = -\left[y \cdot \log \sigma(wx) + (1 - y) \cdot \log(1 - \sigma(wx))\right], \tag{3}$$

where $\sigma(s)$ is the logistic function and w are the parameters of the model. Compute the gradient of the above loss function with respect to the parameter vector w. Show all the steps of the derivation.

Practicum

See the accompanying Python notebook.

Question P1: Linear Regression (20 points)

Question P2: Gradient Descent (10 points)

Question P3: Logistic Regression (40 points)