**Machine Learning: Assignment 5**

Due on Thur, Mar 23, 2023 at 11:00am

Instructor: Dr. Anh Nguyen Note: Late assignments will NOT be graded at all unless you have been granted extensions in advance (see the policy in syllabus).

There are 3 problems for a total of 100 points.

Submission:

* For Problem 0, generate A5\_problem0.ipynb.
* Please write the answers to Problem 2, 3 in a typeset format (i.e., all your answers must be typed up using Word or Latex) and then generate a single PDF named A5\_problem1-2.pdf. Handwritten work will not be graded.
* Please zip up both 2 files (A5\_problem0.ipynb, A5\_problem1-2.pdf) into a single zip file and upload that to Canvas.

**Problem 0**

40 points Run all the cells in this Google Colab. Please read all code and the instructions carefully and complete all the lines that says:

## <--- EDIT THIS LINE

Submission Download your completed Google Colab as ipynb and rename it as A5\_problem0.ipynb

**Problem 1  
Linear Regression [30 pts]**

Suppose that, , where

a) [10 pts] Write down an expression for .

b) [10 pts] Assume you are given a set of training observations for . Write down the conditional log-likelihood of this training data. Drop any constants that do not depend on the parameters , or .

c) [10 pts] Based on your answer, show that finding the MLE of that conditional log- likelihood is equivalent to minimizing least squares,

MLE of conditional log-likelihood

Minimizing least squares

You can see pretty clearly that the MLE of the conditional log likelihood is equivalent to minimizing least squares, once constants are removed the form is identical.

**Problem 2**

**Regularization [30 pts]**

a) [15 pts] Find the partial derivative of the regularized least squares problem:

with respect to , and . Although there is a closed-form solution to this problem, there are situations in practice where we solve this via gradient descent. Write down the gradient descent update rules for , and .

Let us denote

The partial derivative of with respect to is:

The partial derivative of with respect to is:

The partial derivative of with respect to is:

The gradient descent update rules for is:

b) [15 pts] Suppose that . Prove that, the MAP estimate of , and with this prior is equivalent to minimizing the above regularized least squares problem with

Hint: Derive the equations for the two optimization problems and show they are equivalent.

Write in summation form

MAP

L2 Regularization

Leaving us with a likelihood function P(D|w) =

And Prior, P(w) =

Thus MAP is

Taking the Log of both sides

Which is exactly what we see in the minimization form,