HW3

CISC648010 - Spring 2022

Due Date: March 4th at 11 PM

1 Problem 1 (Outer Product of Two Vectors) 25 pts

a. Let $x_1 = \begin{bmatrix} -1 \\ 1 \\ -2 \end{bmatrix}$. Calculate matrix $A_1 = x_1 \cdot x_1^T$. Note that A_1 is called the outer product of x_1.

b. Is matrix A_1 positive semi definite? Is matrix A_1 positive definite?

c. Let $x_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$. Calculate $A_2 = x_2 \cdot x_2^T$.

d. Define $A = A_1 + A_2$. Is matrix A positive semi definite? Is matrix A positive definite?

e. Define $B = A + 0.2I$, where matrix I is a 3 by 3 Identity matrix. Is B positive definite? why?

Remark This problem may help you to understand why the loss function of regularized logsitic regression (J_λ) is strictly convex.

2 Problem 2 (Naive Bayse Spam filtering) 15 pts

Last week, you implemented LDA for Sapm Filtering. This week, you use Naive Bayes to do that.

Download **NB.py** from Canvas. This code loads spambase dataset. In this code x is $n \times d$ where $n = 4601$ and $d = 57$. The different features correspond to different properties of an email, such as frequency with which

certain characters appear y is a vector of labels indicating spam (label 1) not spam (label 0). For the detailed description of the dataset, visit the UCI Machine Learning Repository or Google Spambase.

Note that features in the spambase dataset are not binary. The provided code processes/manipulates the data to make the features binary. It also splits the data into training and test dataset.

In this problem, you should train a Naive Bayes classifier for the problem of spam detection, using the training data. Use the test data to report the test error. Upload your code on Canvas as a single file named **NB_<lastname>.py**. Please also report the test error.

Notes: The base codes have been written using python 3. If you get an error when you run problem2.py or problem1.py, you may be using a wrong version of python.

If you do not know which environment you should use for writing a python code, you can start with google colab. If you need help with google colab, please check here: https://www.youtube.com/watch?v=i-HnvsehuSw