CS 6220 Data Mining — Assignment 5
Due: February 1, 2023(100 points)

YOUR NAME
YOUR GIT USERNAME
YOUR E-MAIL

Naïve Bayes, Bayes Rules

The original performance of acoustic classification for Parkinsons Disease leverages speech recordings from controlled subject responses from variety of questions. The task in the competition was to detect whether or not a person X had Parkinsons disease from a sampling of data. As of 2018, the state of the art classifiers have achieved 90% correct classification on a held out dataset, both for subjects who had Parkinsons and those who did not (at equal rates). So, when classifier Y sees person X, it works correctly 90% of the time.

Question 1 [25 pts total]

Let's say that we run a clinic. This clinic leverages this classifier, which has 90% accuracy. Also, let us say that we know that our current patient load is that 10% of the population have Parkinsons and 90% of the population do not. Let's also say that we're seeing patient X, and the classification algorithm has detected that they have Parkinson's disease. What's the probability that indeed X has Parkinson's disease?

Come up with the numerical solution, and show your written work.

Gradient Descent - Logistic Regression

Yann LeCun, the Director of Facebook Research and one of the fathers of deep learning neural networks, got his start with the MNIST Dataset, which is widely regarded as the "Hello World" of Computer Vision. The MNIST dataset can be downloaded at LeCun's website.
Logistic regression's cost function is typically binary cross-entropy.

$$\mathcal{L}(W, b) = - \sum_i y_i \log h_{W,b}(\mathbf{x}) + (1 - y_i) \log (1 - h_{W,b}(\mathbf{x}))$$

where

$$h_{W,b}(\mathbf{x}) = \sigma \left(W^T \mathbf{x} + b \right)$$

and

$$\sigma(z) = \frac{1}{1 - \exp(-z)}$$

Question 2a.)

Prove that the derivative of the sigmoid function is $\nabla_z \sigma(z) = \sigma(z) \left(1 - \sigma(z) \right)$.

Question 2b.)

Given the cost function defined above, let's assign the estimate

- The gradient of $\mathcal{L}$ with respect to W: $\nabla_W \mathcal{L} = (\mathbf{y} - h_{W,b}(\mathbf{x})) \cdot \mathbf{x}^T$

- The gradient of $\mathcal{L}$ with respect to b: $\nabla_b \mathcal{L} = (\mathbf{y} - h_{W,b}(\mathbf{x}))$