



NORTHEASTERN UNIVERSITY, KHOURY COLLEGE OF COMPUTER SCIENCE

---

## 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.

1. 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.