# CS 542 (Fall 2022) Written Assignment 1
# Bayes' Theorem and Naïve Bayes Classification

Due September 15, 2022

## 1 Bayes' Theorem

You get an email. You know that 90% of your email is legitimate ($L$) while 10% is spam ($S$).

a. Assume the following probabilities:

- The probability that an email contains the word "Bitcoin" ($B$) if it is spam is 96%.
- The probability that an email contains the word "Bitcoin" if it is legitimate is 5%.

What is the probability that the your new email is spam given that it contains the word "Bitcoin"? **Show your work!**

b. Assume the following probabilities:

- The probability that an email contains the word "Covid" ($C$) if it is spam is 50%.
- The probability that an email contains the word "Covid" if it is legitimate is 12%.

What is the probability that the your new email is legitimate given that it contains the word "Covid"? **Show your work!**

## 2 Naïve Bayes

The following problem is from the Jurafsky and Martin book, Exercise 4.2, reproduced below.

Given the following short movie reviews, each labeled with a genre, either comedy or action:

| document | class |
|---|---|
| fly fast shoot love | action |
| fun couple love love | comedy |
| fast furious shoot | action |
| couple fly fast fun fun | comedy |
| furious shoot shoot fun | action |

and a new document D:

fast couple shoot fly

compute the most likely class for D. Assume a naive Bayes classifier and use add-1 smoothing for the likelihoods.

**Show your work!** In particular, show all of the probability distributions involved in the model (namely, $P$(class) and $P$(feature|class)) and all of the steps used to calculate them. Create (conditional) probability tables such as those shown below.

| class | $P$(class) |
|---|---|
| action | |
| comedy | |

| $P$(feature\|class) | | feature | | | |
|---|---|---|---|---|---|
| | | fast | couple | shoot | fly |
| class | action | | | | |
| | comedy | | | | |

Perform Laplace Smoothing to account for words that do not appear in one class.

## Submission Instructions

Please submit your solutions (in PDF format - printed and scanned images are OK) to the drop box on Canvas.