CSDS 391 Introduction to Artificial Intelligence
Written Assignment 5 (“W5”)

Total Points: 100

Remember: name and case ID, answers concise, neat, and legible. Submit electronically via canvas. You can scan your hand-written assignment, but make sure all your answers are legible. Your file must be a .pdf file (.doc or .txt files are not allowed), and the filename should have the following format: `W5_yourCaseID.pdf`.

Note: Some of the questions below ask you to make plots and/or write simple programs. This might be more convenient to do in a language with a good plotting library such as Matlab or python using matplotlib.

Q1. Bernoulli trials and bias beliefs

Recall the binomial distribution describing the likelihood of getting y heads for n flips:

$$p(y|\theta, n) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$

where θ is the probability of heads.

a) Using the fact

$$\int_0^1 p(y|\theta, n) d\theta = \frac{1}{1+n}$$

derive the posterior distribution for θ assuming a uniform prior. (10 P.)

b) Plot the likelihood for $n = 4$ and $\theta = 3/4$. Make sure your plot includes $y = 0$. (10 P.)

c) Plot the posterior distribution of θ after *each* of the following coin flips: head, head, tail, head. You should have four plots total. (10 P.)

Q2. After R&N 20.1 Bags O’ Surprise

The data used for Figure 20.1 on page 804 can be viewed as being generated by h_5.

a) For each of the other four hypotheses, write code to generate a data set of length 100 and plot the corresponding graphs for $P(h_i|d_1, \ldots, d_N)$ and $P(D_{N+1} = \text{lime}|d_1, \ldots, d_N)$. The plots should follow the format of Figure 20.1. Comment on your results. (20 P.)

b) What is the mathematical expression for how many candies you need to unwrap before you are more 90% sure which type of bag you have? (10 P.)

c) Make a plot that illustrates the reduction in variabilty of the curves for the posterior probability for each type of bag by averaging each curve obtained from multiple datasets. (20 P.)

Q3. k-means Clustering

In k-means clustering, $\boldsymbol{\mu}_k$ is the vector mean of the k^{th} cluster. Assume the data vectors have I dimensions, so $\boldsymbol{\mu}_k$ is a (column) vector $[\mu_1, \ldots, \mu_I]_k^T$, where the symbol ^{T} indicates vector transpose.

Derive update rule for $\boldsymbol{\mu}_k$ using the objective function

$$D = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{n,k} \|\boldsymbol{x}_n - \boldsymbol{\mu}_k\|^2$$

where $\boldsymbol{x}_n$ is the n^{th} data vector, $r_{n,k}$ is 1 if $\boldsymbol{x}_n$ is in the k^{th} class and 0 otherwise, and $\|\boldsymbol{x}\|^2 = \boldsymbol{x}^T\boldsymbol{x} = \sum_i x_i x_i = \sum_i x_i^2$. The update rule is derived by computing the gradient for each element of the k^{th} mean and solving for the value where the gradient is zero. Express your answer first in scalar form for $\mu_{k,i}$ and in vector form for $\boldsymbol{\mu}_k$. (20 P.)