EECS 391 Introduction to Artificial Intelligence
Fall 2017, Written Assignment 5 ("W5")

Due: Tue Dec 5 in class

Total Points: 100 + 25 extra credit

Remember: name and case ID, stapled, answers concise, neat, and legible. Submit in class on the due date.

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, Mathematica, or python using matplotlib. Submit your code via Canvas, but turn in the homework writeup in class.

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. (5 P.)

b) Plot the likelihood for $n = 4$ and $\theta = 3/4$. Make sure your plot includes $y = 0$. (5 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. (15 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? (5 P.)

c) Make a plot that illustrates the reduction in uncertainty for each type of bag by averaging multiple random data sets. (15 P.)

Q3. Classification with Gaussian Mixture Models

Suppose you have a random variable x which is drawn from one of two classes C_1 and C_2. Each class follows a Gaussian distribution with means μ_1 and μ_2 (assume $\mu_1 < \mu_2$) and variances σ_1 and σ_2. Assume that the prior probability of C_1 is twice that of C_2.

a) What is the expression for the probability of x, i.e. $p(x)$, when the class is unknown? (5 P.)

b) What is the expression for the probability of total error in this model assuming that the decision boundary is at $x = \theta$? (10 P.)

c) Derive an expression for the value of the decision boundary θ that minimizes the probability of misclassification. (10 P.)

Q4. 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.

a) 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.)

b) Extra credit. Write a program that implements the k-means clustering algorithm on the iris data set. Plot the results of the learning process by showing the initial, intermediate, and converged cluster centers for $k = 2$ and $k = 3$. (+25 P. bonus)