ECE M146 Homework 1
Introduction to Machine Learning
Instructor: Lara Dolecek
TA: Zehui (Alex) Chen, Ruiyi (John) Wu

Please upload your homework to Gradescope by April 9, 11:59 pm.
You can access Gradescope directly or using the link provided on CCLE.
You may type your homework or scan your handwritten version. Make sure all the work is discernible.

1. Assume that there are two urns. The first urn contains 4 red balls, 3 blue balls, and 3 white balls. The second urn contains 2 red balls, 4 blue balls, and 4 white balls. You randomly select an urn and take two balls from the urn. The probability that you pick the first urn is 40%. What is the probability that
   (a) the two balls are red?
   (b) the second ball is blue?
   (c) the second ball is blue given that the first ball is red?
2. Suppose 6 identical dice each with faces numbered 1 through 6 are tossed at the same time. What is the probability of the event "the result of the outcome is such that three different numbers each appear twice?"
3. In a bolt factory, machines A, B, C manifacture, respectively 25, 35 and 40 per cent of the total. Of their product 5, 4, and 2 per cent are defective bolts. A bolt is drawn at random from the produce and is found defective. What are the probabilities that it was manufactured by machines A, B and C?
4. Let X and Y be discrete random variables. Let $\mathbb{E}[X]$ and $var[X]$ be the expected value and variance, respectively, of a random variable X.
   (a) Show that $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.
   (b) If X and Y are independent, show that $var[X+Y] = var[X] + var[Y]$.
5. Suppose that you are waiting at a bus stop. The waiting time until a bus arrives is T where T is an exponentially distributed random variable with parameter λ i.e. $P(T \leq t) = 1 - e^{-\lambda * t}, \forall t \geq 0$.
   (a) Given that you have already waited r seconds, what is the probability that the bus will not arrive within d more seconds?
   (b) What is the average waiting time for the bus i.e. the expected value of T? Hint: Recall that one way to solve $\int udv$ is by integration by parts.

6. In this exercise, you will implement the perceptron algorithm in MATLAB. You will be provided with 3 datasets: *data1.csv*, *data2.csv*, and *data3.csv*. Each dataset will have three columns. The first two columns are the attributes of the datapoint and the third column is the label for each datapoint. The attributes have been normalized so that $\|x\| \leq 1$. Each label is either 1 or -1.

   (a) Plot all datasets. Which datasets are linearly separable?

   (b) Implement the perceptron algorithm as shown in chapter 4 of *A Course in Machine Learning*. To allow for the same results, initialize the hyperplane parameters as 0, iterate through data points in the order provided. Set the maximum iteration number to 1000. For each dataset, provide the hyperplane parameters that are learned by the perceptron algorithm (w and b) and report the total number of updates performed (u). In addition, for each data set, provide a plot that shows both the data and the decision boundary, i.e., the line defined by $w^T x + b = 0$. Based on the total number of updates performed, comment on the convergence of perceptron algorithm for each data set.

   (c) Now, you will compare the rate of convergence for the linearly separable datasets. Recall that the margin $\gamma_{w,b}$ is the distance between the hyperplane defined by $\{w, b\}$ and the nearest point of a set. The margin γ of a set is the largest $\gamma_{w,b}$ for all hyperplanes $\{w, b\}$ that separate the set. As shown in lecture, the number of updates needed to converge is upper bounded by $\frac{1}{\gamma^2}$. Unfortunately, we currently do not have the tools to find γ (will be discussed when the course reaches SVMs). Fortunately, we can use the hyperplane (defined by w and b) found by the perceptron algorithm to get an lower bound on the margin since by definition $\gamma_{w,b} \leq \gamma$ which implies that $\frac{1}{\gamma^2} \leq \frac{1}{\gamma_{w,b}^2}$.

   For each linearly separable dataset, calculate the margin $\gamma_{w,b}$ using the learned parameters and compare the upper bound ($\frac{1}{\gamma_{w,b}^2}$) to the number of updates that you actually had.