ECE M146 | Homework 1
Introduction to Machine Learning | Monday, March 29, 2021
Instructor: Lara Dolecek | Due: Monday, April 5, 2021
TA: Zehui (Alex) Chen | chen1046@ucla.edu

Please upload your homework to Gradescope by April 5, 4:00 pm.
Please submit a single PDF directly on Gradescope
You may type your homework or scan your handwritten version. Make sure all the work is discernible.

1. Let
$$\boldsymbol{x} = \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}, \boldsymbol{y} = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}.$$
Also let θ be the angle between $\boldsymbol{x}$ and $\boldsymbol{y}$. Calculate the following expressions:
   (a) $\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{y}$.
   (b) $\|\boldsymbol{x}\|_1$ and $\|\boldsymbol{y}\|_1$.
   (c) $\|\boldsymbol{x}\|_2$ and $\|\boldsymbol{y}\|_2$.
   (d) $cos(\theta)$.

2. In a bolt factory machines A, B, C manufacture, respectively 25, 35 and 40 percent of the total products. Of their product 5, 4, and 2 per cent are defective bolts. A bolt is drawn at random from the products and is found defective. What are the probabilities that it was manufactured by machines A, B and C?

3. 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 $var[X + Y] = var[X] + var[Y]$.

4. (a) Let $x_1, x_2, \cdots, x_n$ be identically distributed random variables. A random vector, $\boldsymbol{x}$, is defined as $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^T$. What is $\mathbb{E}[\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}]$ in terms of $\mathbb{E}[\boldsymbol{x}]$, given that $\boldsymbol{A}$ and $\boldsymbol{b}$ are deterministic?
   (b) Let
$$\boldsymbol{cov}(\boldsymbol{x}) = \mathbb{E}[(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])^T].$$
   What is $\boldsymbol{cov}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$ in terms of $\boldsymbol{cov}(\boldsymbol{x})$, given that $\boldsymbol{A}$ and $\boldsymbol{b}$ are deterministic?
   (c) Let $\boldsymbol{x}$ be a random vector that follow a multivariate Gaussian distribution defined by its mean $\mathbb{E}[\boldsymbol{x}]$ and covariance matrix $\boldsymbol{cov}(\boldsymbol{x}) = \mathbb{E}[(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])^T]$. What is the distribution of $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$, given that $\boldsymbol{A}$ and $\boldsymbol{b}$ are deterministic?

5. You will implement the perceptron algorithm in this problem. You are provided with 2 datasets: *UCLA_EE_grad_2030.csv* and *UCLA_EE_grad_2031.csv*. These datasets contain UCLA graduate student admission data in 2030 and 2031, respectively. Each dataset will have three columns. The first two columns contain two features for the applicants. The first column represents the GPA of applicants and the second column represents the normalized GRE score of applicants. The third column contains labels denoting whether the applicant is admitted into UCLA or not. Each label is either 1 or -1 where 1 denotes a student being admitted and -1 denotes a student not being admitted. Throughout this quarter, we will use these two datasets repetitively to text different learned algorithms.

   (a) Plot all the 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 learned 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) Recall that the empirical margin $\gamma_{w,b}$ is the distance between the hyperplane defined by $\{w, b\}$ and the nearest point of a set. Calculate $\gamma_{w,b}$ for the linearly separable dataset with your learned parameters.