CS 559: Machine Learning

Homework Assignment 4

Due Date: Wednesday 6:30 PM, April 27, 2022 **Total: 100 points**

Problem 1. K-means (15 points)

Consider the following six points (with (x, y) representing location in a 2D space) and let us try to group them into three clusters.

$$A_1 = (2,10), A_2 = (2,5), A_3 = (8,4), B_1 = (5,8), C_1 = (1,2), C_2 = (4,9)$$

The distance function is Euclidean distance. Suppose initially we assign A_1, B_1 and C_2 as the center of each cluster. Using the k-means algorithm to answer the following questions.

1) Cluster assignment of each data point after the first iteration.
2) Centroid after the first iteration.

Problem 2. Expectation Maximization (EM) for GMM (40 points)

In this question you will implement the EM algorithm for Gaussian Mixture Models. A good read on gaussian mixture EM can be found at this link: https://www.ics.uci.edu/~smyth/courses/cs274/notes/EMnotes.pdf. A two-dimensional dataset for this problem can be downloaded in canvas. For this problem,

- n is the number of training points
- f is the number of features
- k is the number of gaussians
- X is an $n \times f$ matrix of training data
- r is an $n \times k$ matrix of membership weights. $r(i, j)$ is the probability that the data sample x_i was generated by gaussian j
- π is a $k \times 1$ vector of mixture weights (gaussian prior probabilities). π_i is the prior probability that any point belongs to cluster i.
- μ is a $k \times f$ matrix containing the means of each gaussian
- Σ is an $f \times f \times k$ tensor of covariance matrices. $\Sigma(:, :, i)$ is the covariance of gaussian i

Questions:

1) **Expectation**: Complete the function $[r] = Expectation(X, k, \pi, \mu, \Sigma)$. This function takes in a set of parameters of a gaussian mixture model, and outputs the membership weights of each data point.
2) **Maximization of Means**: Complete the function $[\mu] = MaximizeMean(X, k, r)$. This function takes in the training data along with the membership weights, and calculates the new maximum likelihood mean for each gaussian.
3) **Maximization of Covariances**: Complete the function $[\Sigma] = MaximizeCovariance(X, k, r, \mu)$. This function takes in the training data along with membership weights and means for each gaussian, and calculates the new maximum likelihood covariance for each gaussian.
4) **Maximization of Mixture Weights**: Complete the function $[\pi] = MaximizeMixtures(k, r)$. This function takes in the membership weights, and calculates the new maximum likelihood mixture weight for each gaussian.

5) **EM**: Put everything together and implement the function $[\pi, \mu, \Sigma] = EM(X, k, \pi_0, \mu_0, \Sigma_0, nIter)$. This function runs the EM algorithm for $nIter$ steps/iterations and returns the parameters of the underlying GMM.
Note: Since this code will call your other functions, make sure that they are correct first. A good way to test your EM function offline is to check that the log likelihood $\log P(X|\pi, \mu, \Sigma)$ is increasing for each iteration of EM. *Make sure to work on the implementation independently.*

Problem 3. Bayesian Classification (15 points)

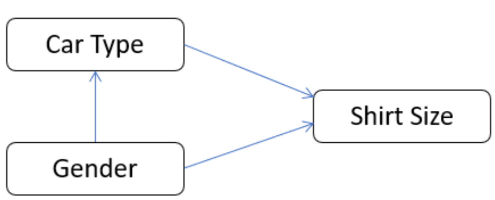
Consider the training examples shown in the following Table for a binary classification problem.

Customer ID	Gender	Car Type	Shirt Size	Class
1	M	Family	Small	C0
2	M	Sports	Medium	C0
3	M	Sports	Medium	C0
4	M	Sports	Large	C0
5	M	Sports	Extra Large	C0
6	M	Sports	Extra Large	C0
7	F	Sports	Small	C0
8	F	Sports	Small	C0
9	F	Sports	Medium	C0
10	F	Luxury	Large	C0
11	M	Family	Large	C1
12	M	Family	Extra Large	C1
13	M	Family	Medium	C1
14	M	Luxury	Extra Large	C1
15	F	Luxury	Small	C1
16	F	Luxury	Small	C1
17	F	Luxury	Medium	C1
18	F	Luxury	Medium	C1
19	F	Luxury	Medium	C1
20	F	Luxury	Large	C1

Now given the 20 training samples, classify the new sample with the following attribute values using Naïve Bayes and Bayesian network, respectively. Provide the detailed calculations.

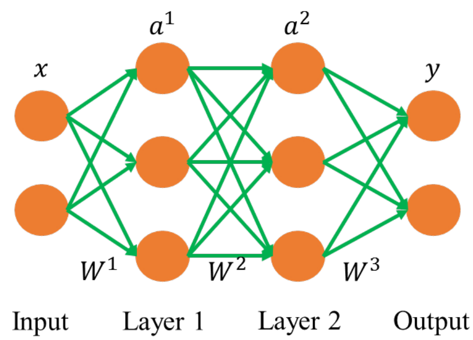
21	Gender=M	Car Type=Family	Shirt Size=Large	?

1) Classify the new test example using **Naïve Bayes** method.
2) Classify the new test example using **Bayesian Network** approach based on the given network structure below.



Problem 4. Neural Networks (30 points)

Consider the following deep feedforward neural network with two hidden layers.



Assumptions:

a) We have zero biases in each layer, i.e., $z^1 = W^1x$, $z^2 = W^2a^1$, $z^3 = W^3a^2$.
b) The activation function for two hidden layers are sigmoid function, i.e., $a^1 = \sigma(z^1)$ and $a^2 = \sigma(z^2)$, where $\sigma(z) = \frac{1}{1+e^{-z}}$.
c) The activation function for the output layer is a SoftMax function, i.e., $y_i = \frac{e^{z_i^3}}{\sum_j e^{z_j^3}}$.
d) The loss function is given by $L = \|y - \hat{y}\|^2 = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2$.

Question:

Using backpropagation method and the computational graph learned in the class to derive $\frac{\partial L}{\partial W^1}$, $\frac{\partial L}{\partial W^2}$, $\frac{\partial L}{\partial W^3}$. Show the details of your work and solutions.