

The objective of this homework is to test your understanding of the module 9 and Machine Learning Lab lecture video. Due date of the homework is: **Monday, November 30, 2020**. You have to upload a scanned version of your solutions on Blackboard. If you don't have a scanner around you, you can use Cam Scanner — a mobile app that scans images in a neat way, as if they're scanned through a copier.

For questions related to the homework, please email the two teaching assistants: (a) Wasim (email: wasim.dipon@my.utsa.edu) and (b) Yuvaraj (email: yuvaraj.munian@my.utsa.edu).

You can send your question to both of them on the same email. Feel free to CC me as well, but that is not necessary.

1. What is meant by Machine Learning and Artificial Intelligence? (5 Marks)
2. What are the different machine learning algorithms? (5 Marks)
3. Andrade's equation has been proposed as a means of modelling the effects of temperature on viscosity. The equation is,

$$\mu = D * e^{B/T_A}$$

Where, μ is the dynamic viscosity

T_A is the absolute temperature (K)

D and B are unknown parameters

- a. Write a MATLAB code function *fitExp()* (save the function as *fitExp.m*) that takes as inputs μ and T and the outputs D and B. Within the function: [Hint: Refer the table] (20 Marks)
 - i. Convert the data so that you can solve this as a linear regression problem and find the best fit straight line. Take into account that T is in Celsius and in the formula, T_A is in kelvin.
 - ii. Produce a plot of the transformed data and the straight line.
 - iii. Now find the values of D and B that best fit the equation to the data. Output the value using *fprintf*.
 - iv. Produce a plot of the original data and the fitted curve.

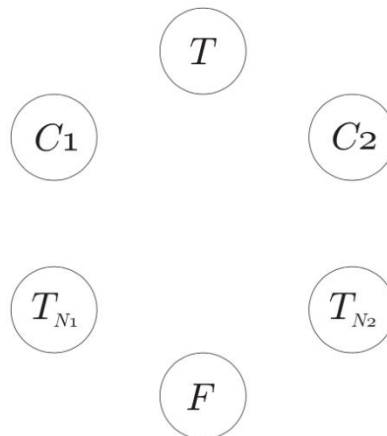
T(°C)	0	5	10	20	30	40
μ	1.787	1.519	1.307	1.002	0.7975	0.6529

[Hint: Recall, K = °C + 273.15]

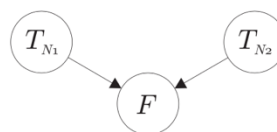
- b. Write a script file called hw6p3.m and create vectors T and μ (μ) in your script using the values from the following table and uses the fitExp() function you created. (20 Marks)
4. We wish to develop a graphical model for the following transportation problem. A transport company is trying to choose between two alternative routes for commuting between Boston and New York. In an experiment, two identical busses leave Boston at the same but otherwise random time, TB . The busses take different routes, arriving at their (common) destination at times $TN1$ and $TN2$.

Transit time for each route depends on the congestion along the route, and the two congestions are unrelated. Let us represent the random delays introduced along the routes by variables $C1$ and $C2$. Finally, let F represent the identity of the bus which reaches New York first. We view F as a random variable that takes values 1 or 2.

- a. Complete the following directed graph (Bayesian network) with edges so that it captures the relationships between the variables in this transportation problem. (20 Marks)



- b. Consider the following directed graph as a possible representation of the independences between the variables $TN1$, $TN2$, and F only:

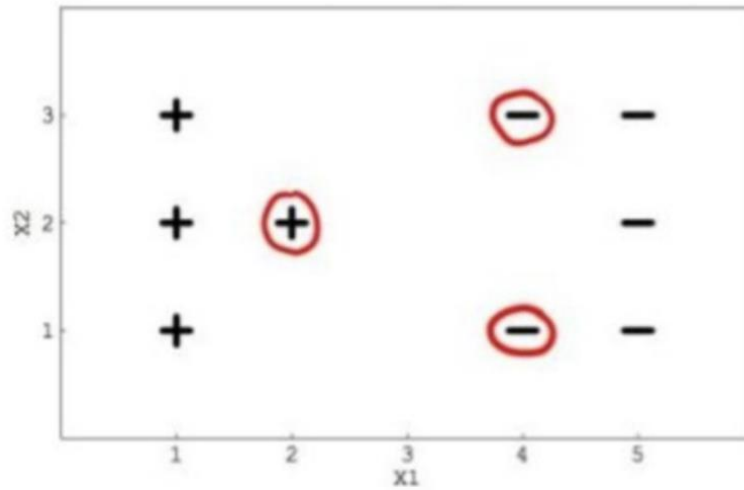


Which of the following factorizations of the joint are consistent with the graph? (15 Marks)

$P(TN1) P(TN2) P(F|TN1, TN2)$ ☐

$P(TN1) P(TN2) P(F|TN1)$ ☐

5. Define Linear Support Vector Machine.
- a. Suppose you are using a linear SVM classifier with 2 class classification problem. Consider the following data in which the points circled red represent support vectors.



Will the decision boundary change if any of the red points are removed. If Yes, give the reason? (15 Marks)