INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

Stamp / Signature of the Invigilator												
EXAMINATION (End Semester)										SEMESTER (Autumn)		
Roll Number										Section		Name
Subject Number	C	S	6	0	0	5	0	Subject Name		*MACHINE LEARNING*		
Department / Center of the Student										Additional sheets		

Instructions and Guidelines to Students Appearing in the Examination

1. Ensure that you have occupied the seat as per the examination schedule.
2. Ensure that you do not have a mobile phone or a similar gadget with you even in switched off mode. Note that loose papers, notes, books should not be in your possession, even if those are irrelevant to the paper you are writing.
3. Data book, codes or any other materials are allowed only under the instruction of the paper-setter.
4. Use of instrument box, pencil box and non-programmable calculator is allowed during the examination. However, exchange of these items is not permitted.
5. Additional sheets, graph papers and relevant tables will be provided on request.
6. Write on both sides of the answer script and do not tear off any page. Report to the invigilator if the answer script has torn page(s).
7. Show the admit card / identity card whenever asked for by the invigilator. It is your responsibility to ensure that your attendance is recorded by the invigilator.
8. You may leave the examination hall for wash room or for drinking water, but not before one hour after the commencement of the examination. Record your absence from the examination hall in the register provided. Smoking and consumption of any kind of beverages is not allowed inside the examination hall.
9. After the completion of the examination, do not leave the seat until the invigilator collects the answer script.
10. During the examination, either inside the examination hall or outside the examination hall, gathering information from any kind of sources or any such attempts, exchange or helping in exchange of information with others or any such attempts will be treated as adopting 'unfair means'. Do not adopt 'unfair means' and do not indulge in unseemly behavior as well.

Violation of any of the instructions may lead to disciplinary action.

Signature of the Student

To be filled in by the examiner

Question Number	1	2	3	4	5	6	7	8	9	10	Total
Marks Obtained											

Marks obtained (in words)	Signature of the Examiner	Signature of the Scrutineer

Instructions: Answer all FIVE questions. Time = 3hrs. Total marks = 5x20 = 100. Write your answers only in the space provided. Show the solution steps as required. The question paper has total 12 pages.

ROUGH WORK

1.(a). Write the objective function for the primal optimization problem for soft margin SVM. Briefly, define the terms. [2]

(b). Write the objective function for the dual optimization problem for soft margin SVM. Briefly, define the terms. [2]

(c). For support vectors X_j in a hard margin SVM, $WX + b = 0$, we have: $|WX + b| =$ ☐ [2]

(d). We are designing a SVM, $WX + b = 0$, suppose (X_j, y_j)'s are the support vectors and α_j's the corresponding Lagrange multipliers, then W = ☐ [2]

(e). We are designing a SVM, $WX + b = 0$, suppose (X_j, y_j)'s are the support vectors and α_j's the corresponding Lagrange multipliers, then $\sum \alpha_j y_j =$ ☐ [2]

(f). Consider the following training data set of (seven) points X's in a plane and their binary class label y's:

X	*(1, 0)*	*(0, 1)*	*(0, -1)*	*(-1, 0)*	*(0, 2)*	*(0, -2)*	*(-2, 0)*
y	*-1*	*-1*	*-1*	*+1*	*+1*	*+1*	*+1*

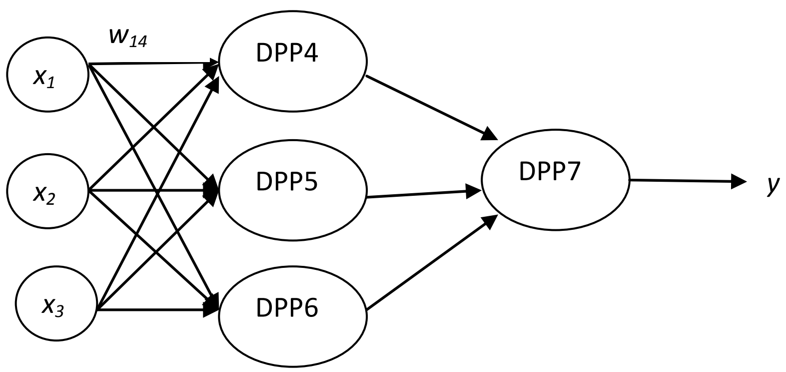
We perform the following non-linear transform of the input vector $X = (x_1, x_2)$ to obtain the transformed feature vector $Z = (z_1, z_2) = (\phi_1(X), \phi_2(X))$, with $\phi_1(X) = x_2^2 - 2x_1 + 3$, $\phi_2(X) = x_1^2 - 2x_2 - 3$.

Write the equation of the optimal separating hyperplane in transformed space Z. Explain your answer. [10]

Equation of optimal separating hyperplane:

Explanation:

2.(a). A dot product perceptron (DPP) has output $y = W{\cdot}X$, where W is the weight vector and X is the input vector. There is no bias input. Consider the network of dot product perceptrons shown below. The input (nodes) are numbered 1, 2, and 3. The other nodes are numbered (5-7) as shown in figure. Weight w_{ij} connects node *i* to node *j*. It can be shown that this network is equivalent to a single dot product perceptron. Draw the equivalent dot product perceptron showing its weights in terms of the original weights. [10]



(b). We want to approximate the function: $t(x) = x^2,\ x \in [2, 4]$, using a perceptron with a single input x and the bias set to 1. The perceptron has a linear activation function. The approximation performance of the network is measured with the following error function: $E = \int_2^4 [t(x) - y(x)]^2\, dx$, where x is the input and $y(x)$ is the corresponding output. Derive the weights of the optimal perceptron for this task. Draw the perceptron showing the weights. [10]

3.(a). Consider a set *D* of six two-dimensional points: *a* = (0, 0), *b* = (5, 0), *c* = (10, 0), *d* = (0, 5), *e* = (5, 8), *and f* = (10, 8). We cluster these points using k-means algorithm with *k* = 3 and Euclidean as the distance metric. Ties are broken in favor of the mean to the left/bottom. A "k-seed set" is a subset of *D* that form the initial means, e.g., *{a, b, c}* is a 3-seed set. A "k-clustering" is a partition of *D* into *k* non-empty subsets, e.g., *{a, d, f} {b, c}, {e}* is a 3-clustering. Naturally, a k-clustering has a corresponding set of *k* means. A k-clustering is called "stable" if an iteration of the k-means algorithm using the corresponding means leaves it unchanged. Fill in the following table about the given 3-clusterings. [10]

3-clustering	*Stable?*	*If stable, list all the k-seed sets that converge to the 3-clustering after some iterations of the k-means algorithm.*
{a, b, e} {c, d} {f}		
{a, b, d} {e, f} {c}		
{a} {d} {b, c, e, f}		

(b) Consider a set of *N* one-dimensional points: $\{2^0, 2^1, 2^2, \ldots, 2^{N-1}\}$, with Euclidean as the distance metric. Draw a schematic of the dendogram for clustering the points using single-linkage algorithm. [5]

(c) Draw a schematic of the dendogram for clustering the above *N* points using complete-linkage algorithm. [5]

4.(a). In the following set of training examples, Feature F1 can take on the values a, b, or c; Feature F2 is Boolean-valued; and Feature F3 is always a real-valued number in [0, 1]. Using the decision stump (a single node decision tree) learner as weak learner, what would the weights be on each of the six training examples *after one round* of AdaBoost? Show your work. [10]

	F1	*F2*	*F3*	*Class*		*F1*	*F2*	*F3*	*Class*
Example 1	a	T	0.2	+	Example 4	b	T	0.6	–
Example 2	b	F	0.5	+	Example 5	a	F	0.1	–
Example 3	b	F	0.9	+	Example 6	a	T	0.7	–

Weight of Example 1: Weight on Example 2: Weight on Example 3:

Weight on Example 4: Weight on Example 5: Weight on Example 6:

(b).The following data set has ten examples each with one real valued feature *x* and one binary output *y*. What is the leave-one-out cross validation error of 1-NN (nearest neighbor) classifier using Euclidean distance on this data set in terms of number of misclassified examples? Explain your answer briefly. [10]

x	-0.1	0.7	1.0	1.6	2.0	2.5	3.2	3.5	4.1	4.9
y	-	+	+	-	+	+	-	-	+	+

5.(a). We have a set of two-dimensional (x_1, x_2) data points whose covariance matrix has eigenvector $\left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]$ with eigenvalue *140*, and eigenvector $\left[\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right]$ with eigenvalue *2*. We want to project the data points to a single dimension (z). Write the expression for z using the original co-ordinates (x_1, x_2) such the projection leads to the minimum squared error loss. Briefly explain your answer. [5]

(b). What is the amount of squared error loss due to this projection? [5]

(c). Define Kullback-Liebler divergence between two distributions. Define the terms used. [5]

(d). When is a feature subset evaluation measure termed as "monotonic"? [5]

------------- **BEST WISHES** -----------

ROUGH WORK

ROUGH WORK

ROUGH WORK

ROUGH WORK