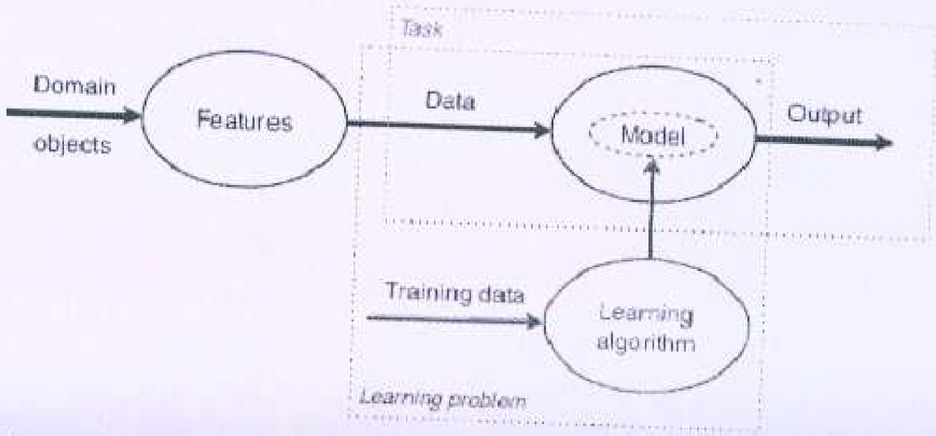
BE/Insem.-62

[redacted]

BE 2012 course In semester examination
Machine Learning (2012)(414455)
Solution scheme

Q 1)a) What is machine learning? Give an overview of machine learning with suitable diagram.
Machine learning is the systematic study of algorithms and systems that improve their knowledge or performance with experience. Or similar to this 1 M



Above diagram with explanation .

b) Explain geometric models and probabilistic models of machine learning in detail. 4 M

Geometric models

Instance space, hyperplane, basic linear classifier, distance, Euclidean distance, nearest-neighbour classifier, linear transformations, support vector machines, etc can be explained. 2.5 M

Probabilistic models

Posterior probability- decision rule, likelihood function, Bayes' rule, prior probability, maximum a posteriori (MAP) decision rule, maximum likelihood (ML) decision rule, posterior odds, prior odds, navie bayes classifier, etc can be explained. 2.5 M

OR

Q 2)a) Explain in detail:

1) Supervised Learning vs Unsupervised learning 2 M
2) Training vs testing 2 M

b) Explain features with suitable example. Give two uses of features. Explain the term discretization of features.

Features with suitable example (may take a dataset) 2 M
Two uses of features-- 'features as splits' and 'features as predictors' 2 M
Discretization-- removal of unnecessary detail contained in real-valued features 2 M

Q 3) a) Define following terms with suitable example:

BE / Insem. - 62

1) Confusion matrix
   In this table, each row refers to actual classes as recorded in the test set, and each column to classes as predicted by the classifier. 2 M
2) False positive rate
   The number of misclassified negatives or false positives as a proportion of the total number of negatives
   $FP/Neg = 1 - tnr$ 2 M
3) True positive rate
   is the proportion of positives correctly classified TP/Pos 2 M

b) What is overfitting? Specify reasons for overfitting.

Overfitting generally occurs when a model is excessively complex, such as having too many parameters relative to the number of observations. 2M

A model that has been overfit will generally have poor predictive performance, as it can exaggerate minor fluctuations in the data. 2M

OR

Q 4) a) What is a contingency table / Confusion Matrix? What is the use of it?

A specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one (in unsupervised learning it is usually called a **matching matrix**). Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. 2 marks

Used in assessing ML algorithms eg accuracy, tnr,tpr,… 2 marks

b) Derive and explain output code matrix for One Vs One and One Vs Rest Scheme for construction of Multi class classifier (for 3 classes). (3 marks each)

Symmetric

$$\begin{pmatrix} +1 & +1 & 0 \\ -1 & 0 & +1 \\ 0 & -1 & -1 \end{pmatrix}$$

asymmetric

$$\begin{pmatrix} +1 & -1 & +1 & -1 & 0 & 0 \\ -1 & +1 & 0 & 0 & +1 & -1 \\ 0 & 0 & -1 & +1 & -1 & +1 \end{pmatrix}$$

Q 5) a) Explain regression using least square method for classification. 6 marks

The *least-squares classifier* learns the decision boundary $\mathbf{w} \cdot \mathbf{x} = t$ with

$$\mathbf{w} = (\mathbf{X}^T\mathbf{X})^{-1}(Pos\,\mu^{\oplus} - Neg\,\mu^{\ominus})$$

We would hence assign class $\hat{y} = \text{sign}(\mathbf{w} \cdot \mathbf{x} - t)$ to instance $\mathbf{x}$, where

BE/Insem. - 62

$$\text{sign}(x) = \begin{cases} +1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$$

OR

May explain with univariate least square classification.

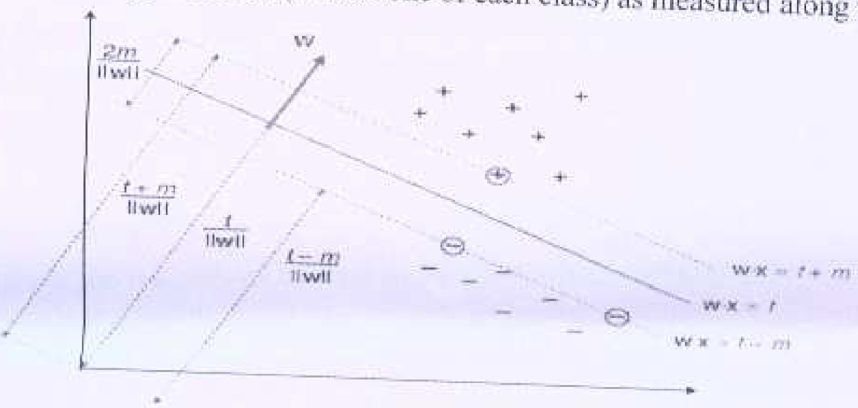
b) What are support vectors and margins? Also explain Soft margin SVM.

Diagram is preferable with explanation

The training examples nearest to the decision boundary are called *support vectors*; as we shall see, the decision boundary of a support vector machine (SVM) is defined as a linear combination of the support vectors. 2M

The *margin* is thus defined as **m**/||**w**||, where m is the distance between the decision boundary and the nearest training instances (at least one of each class) as measured along **w**. 1 M

1 M



Soft margin SVM

2 M

$$\mathbf{w}^*, t^* = \underset{\mathbf{w},t}{\arg\min}\ \frac{1}{2}||\mathbf{w}||^2$$

subject to $y_i(\mathbf{w}\cdot\mathbf{x}_i - t) \geq 1$ and $, 1 \leq i \leq n$

The complexity parameter C can be used to adjust the number and severity of allowed margin violations. Training an SVM involves solving a large quadratic optimisation problem and is usually best left to a dedicated numerical solver

OR

Q 6) a) Explain Perceptron training algorithm for linear classification.

A linear classifier that will achieve perfect separation on linearly separable data is the *perceptron*, originally proposed as a simple neural network. The perceptron iterates over the training set, updating the weight vector every time it encounters an incorrectly classified example.

Explanation of following algorithm 1M

5M

BE/Insem.- 62

Algorithm 7.1: Perceptron(D, η) – train a perceptron for linear classification.

Input : labelled training data D in homogeneous coordinates;
learning rate η.

Output : weight vector $\mathbf{w}$ defining classifier $\hat{y} = \text{sign}(\mathbf{w} \cdot \mathbf{x})$.

```
w ← 0 ;                          // Other initialisations of the weight vector are possible
converged ← false;
while converged = false do
    converged ← true;
    for i = 1 to |D| do
        if y_i w · x_i ≤ 0           // i.e., ŷ_i ≠ y_i
        then
            w ← w + η y_i x_i ;
            converged ← false;       // We changed w so haven't converged yet
        end
    end
end
```

b) What is multivariate regression? Explain its equation using homogeneous coordinates.

Define multivariate regression 1 M

Equation— should be given with explanation 3M

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\mathbf{y} = \mathbf{X}^{\circ}\mathbf{w} + \epsilon$$

$$\hat{\mathbf{w}} = \mathbf{S}^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

$$\hat{\mathbf{w}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$