MACHINE LEARNING ASSIGNMENT

Questions	Options
1	A
2	A
3	B
4	B
5	C
6	B
7	D
8	D
9	A
10	B
11	A
12	A and B

13. Explain the term regularization?

Regularization is an important concept in machine learning . It helps to prevent overfitting of models. At times machine learning models works well with training data but the opposite with test data .It implies that model is having noise in output and hence its overfitted.This problem is solve with regularization techniques.The models may be overfitted or underfitted.
The technique simply allows us to retain all features by reducing the magnitudes of variables.
It basically reduces or regularize the coefficient of features towards zero.
For example
$Y = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_n x_n$
Y= value to be predicted
  $x_1, x_2, \ldots x_n$ =features that decide value of Y
w_0=Bias
$w_1, w_2, \ldots w_n$=weights attached to $x_1, x_2, \ldots x_n$
For accurate predictions we optimize bias and weights.
Therefore we need a loss function and find optimized parameters using gradient descent algorithms and its variance

14. Which particular algorithms are used for regularization?
Regularization algorithms are as follows:

1. LASSO
2. RIDGE

LASSO:It stands for Least Absolute Shrinkage and Selection Operator.It is also called L1 regularization techniques.The equation is given by

$$\text{Loss} = \sum_{j=1}^{m} \left(Yi - Wo - \sum_{i=1}^{n} Wi\, Xji \right)^2 + \lambda \sum_{i=1}^{n} |Wi|$$

Its different from Ridge as it uses absolute weight value for normalization . λ is the tuning parameter and it needs to be tuned using cross validation data set .As loss function only has to consider absolute weights optimization algorithm penalize higher weight values.In LASSO we eliminate parameters having less important features and sets respective weight values to zero .Thus LASSO performs feature selection along with regularization .

RIDGE: It is also known as L2 regularization.It is mainly used for multiple independent variables.Here we add the sum of weights square to a loss function and thus create new loss function

$$\text{Loss} = \sum_{j=1}^{m} \left(Yi - Wo - \sum_{i=1}^{n} Wi\, Xji \right)^2 + \lambda \sum_{i=1}^{n} Wi^2$$

Here normalized weights are in form of squares .λ is the parameter that needs to be tuned using a cross validation data set. If $\lambda=0$ it returns residual sum of squares which we choose initially. If λ is very high ,loss will ignore core loss function and minimize weights square and eventually end up taking parameters value to zero . Now the parameters are learned using a modified loss function . If we want to minimize the function parameters needs to be as small as possible ,therefore L2 norm prevents weights from rising too high.

5. Explain the term error present in linear regression equation?

          An error term represents the margin of error within a statistical model; it refers to the sum of the deviations within the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results. The regression line is used as a point of analysis when attempting to determine the correlation between one independent variable and one dependent variable.
An error term states that the model is not completely accurate and results are different in practical applications. For example,

$Y=\alpha X+\beta\rho+\epsilon$

where:

α,β = Constant parameters

X,ρ = Independent variables

ϵ = Error term

When the actual Y differs from the expected or predicted Y in the model, then the error term is not equal to Zero, which means there are other factors that influence the value of Y.