MACHINE LEARNING
Work sheet-1

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

Q13 to Q15

Q13. Explain the term regularization?

Ans) In mathematics, statistics, finance, computer science, particularly in machine learning and inverse problems, regularization is the process of adding information in order to solve an ill-posed problem or to prevent overfitting.

Regularization can be applied to objective functions in ill-posed optimization problems. The regularization term, or penalty, imposes a cost on the optimization function to make the optimal solution unique.

Independent of the problem or model, there is always a data term, that corresponds to a likelihood of the measurement and a regularization term that corresponds to a prior. By combining both using Bayesian statistics, one can compute a posterior, that includes both information sources and therefore stabilizes the estimation process. By trading off both objectives, one chose to be more addictive to the data or to enforce generalization (to prevent overfitting). There is a whole research branch dealing with all possible regularizations. The work flow usually is, that one tries a specific regularization and then figures out the probability density that corresponds to that regularization to justify the choice. It can also be physically motivated by common sense or intuition, which is more difficult.

In machine learning, the data term corresponds to the training data and the regularization is either the choice of the model or modifications to the algorithm.

It is always intended to reduce the generalization error, i.e., the error score with the trained model on the evaluation set and not the training data. one of the earliest uses of regularization is related to the method of least squares.

In general, regularization means to make things regular or acceptable. This is exactly why we use it for applied machine learning. In the context of machine learning, regularization is the process which regularizes or shrinks the coefficients towards zero. In simple words, regularization discourages learning a more complex or flexible model, to prevent overfitting.

Let X be the set of all possible data points in a d-dimensional space, where each dimension represents one attribute of a data point (which will be used as the input features of the classification model).

We assume there is a predefined set of k classes for data points in X. The objective is to find the relation between each data point and the classes as a classification function $f : X \longrightarrow Y$.

The output reflects how f classifies each input into different classes. Each element of an output $y \in Y$ represents the probability that the input belongs to its corresponding class.

Let Pr (X, Y) represent the underlying probability distribution of all data points in the universe $X \times Y$, where X and Y are random variables for the features and the classes of data points, respectively. The objective of a machine learning algorithm is to find a classification model f that accurately represents this distribution and maps each point in X to its correct class in Y. We assume we have a lower-bounded real-valued loss function l(f (x),y) that, for each data point (x,y), measures the difference between y and the model's prediction f (x). The machine learning objective is to find a function f that minimizes the expected loss:

$L(f) = E_{(x,y)\sim Pr(X,Y)} [l(f(x),y)]$ (1)

We can estimate the probability function Pr(X, Y) using samples drawn from it. These samples construct the training set $D \subset X \times Y$. Instead of minimizing (1), machine learning algorithms minimize the expected empirical loss of the model over its training set D. LD (f) = 1 |D| X (x,y)∈D l(f (x),y) (2) We can now state the optimization problem of learning a classification model as the following: min f LD (f) + λ R(f) (3) where R(f) is a regularization function. The function R(f) is designed to prevent the model from overfitting to its training dataset. For example, the regularization loss (penalty) increases as the parameters of the function f grow arbitrarily large or co-adapt themselves to fit the particular dataset D while minimizing LD (f). If a model overfits, it obtains a small loss on its training data, but fails to achieve a similar loss value on other data points. By avoiding overfitting, models can generalize better to all data samples drawn from Pr(X, Y). The regularization factor λ controls the balance between the classification loss function and the regularization. For solving the optimization problem, especially for nonconvex loss functions for complex models such as deep neural networks, the commonly used method is the stochastic gradient descent algorithm. This is an iterative algorithm where in each epoch of training, it selects a small subset (mini-batch) of the training data and updates the model (parameters) towards reducing the loss over the mini-batch. After many epochs of training, the algorithm converges to a local minimum of the loss function.

Q15. Explain the term error present in linear regression equation?

Ans) An error term is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables. As a result of this incomplete relationship, the error term is the amount at which the equation may differ during empirical analysis.

The error term is also known as the residual, disturbance, or remainder term, and is variously represented in models by the letters *e,* ε, or u.

An error term appears in a statistical model, like a regression model, to indicate the uncertainty in the model. The error term is a residual variable that accounts for a lack of perfect goodness of fit. Heteroskedastic refers to a condition in which the variance of the residual term, or error term, in a regression model varies widely. An error term essentially means that the model is not completely accurate and results in differing results during real-world applications.

Within a linear regression model tracking a stock's price over time, the error term is the difference between the expected price at a particular time and the price that was actually observed. In instances where the price is exactly what was anticipated at a particular time, the price will fall on the trend line and the error term will be zero.

Points that do not fall directly on the trend line exhibit the fact that the dependent variable, in this case, the price, is influenced by more than just the independent variable, representing the passage of time. The error term stands for any influence being exerted on the price variable, such as changes in market sentiment.

The two data points with the greatest distance from the trend line should be an equal distance from the trend line, representing the largest margin of error.

If a model is heteroskedastic, a common problem in interpreting statistical models correctly, it refers to a condition in which the variance of the error term in a regression model varies widely.