Machine Learning Assignment

1) A)

2) A)

3) B)

4) C)

5) C)

6) D)

7) D)

8) D)

9) A)

10) B)

11) B)

12) A) and B)

13) Regularization is the process which regularizes or shrinks the coefficients towards zero. As regularization discourages learning a more complex or flexible model, to prevent overfitting. The basic idea is to penalize the complex models that is adding a complexity term that would give a bigger loss for complex models.

14) Regularization Techniques

There are two main regularization techniques, namely Ridge Regression and Lasso Regression. They both differ in the way they assign a penalty to the coefficients.

Ridge Regression (L2 Regularization)

This regularization technique performs L2 regularization. It modifies the RSS by adding the penalty (shrinkage quantity) equivalent to the square of the magnitude of coefficients. $\sum_{j=1}^{m} (Y_i-W_0-\sum_{i=1}^{n} W_i X_{ji})^2+ \alpha\sum_{i=1}^{n} W_i^2=RSS+ \alpha\sum_{i=1}^{n} W_i^2$

Now, the coefficients are estimated using this modified loss function.

In the above equation, you may have noticed the parameter α (alpha) along with shrinkage quantity. This is called a tuning parameter that decides how much we want to penalize our model. In other terms, tuning parameter balances the amount of emphasis given to minimizing RSS vs minimizing the sum of the square of coefficients.

Let’s see how the value of α alpha affects the estimates produced by ridge regression.

When $\alpha=0$, the penalty term has no effect. It means it returns the residual sum of the square as loss function which we choose initially i.e. we will get the same coefficients as simple linear regression.

When $\alpha=\infty$, the ridge regression coefficient will be zero because the modified loss function will ignore the core loss function and minimize coefficients square and eventually end up taking the parameter's value as 0.

When $0<\alpha<\infty$, for simple linear regression, the ridge regression coefficient will be somewhere between 0 and 1.

That's the reason for selecting a good value of α (alpha) is critical. The coefficient methods produced by ridge regression regularization technique are also known as the L2 norm.

Lasso Regression (L1 Regularization)

This regularization technique performs L1 regularization. It modifies the RSS by adding the penalty (shrinkage quantity) equivalent to the sum of the absolute value of coefficients.
$\sum_{j=1}^{m} (Y_i-W_0-\sum_{i=1}^{n} W_i X_{ji})^2+ \alpha\sum_{i=1}^{n} |W_i |=RSS+ \alpha\sum_{i=1}^{n} |W_i |$
Now, the coefficients are estimated using this modified loss function.

Lasso Regression is different from ridge regression as it uses absolute coefficient values for normalization.

As loss function only considers absolute coefficients (weights), the optimization algorithm will penalize high coefficients. This is known as the L1 norm.

Here, α (alpha) is again a tuning parameter, works like that of ridge regression and provides a tradeoff between balancing RS magnitude of coefficients.

Like ridge regression, α (alpha) in lasso regression can take various values as follows:

When $\alpha=0$, we will get the same coefficients as simple linear regression.

When $\alpha=\infty$, the lasso regression coefficient will be zero.

When $0<\alpha<\infty$, for simple linear regression, the lasso regression coefficient will be somewhere between 0 and 1.

This is all about Regularization and its 2 methods used mostly.

15) Errors are the basically the gap between the best fit points and the points that are present above and below of linear regression line. To mitigate or remove the gap that is known as error we use method like least square method that makes all the points on linear regression line with its equation Y=mx +c we can calculate all the terms such as intercept, slope of line and further we can have points that will close the error or solve the error or gapping between values as more the high and low values the data will be more disturbed and Variance will be high.