ASSIGNMENT – 6

MACHINE LEARNING

Q10 to Q15 are subjective answer type questions, Answer them briefly.

10. Explain how does the adjusted R-squared penalizes the presence of unnecessary predictors in the model?

Ans. The adjusted R-squared is a statistical measure used to evaluate the goodness of fit of a linear regression model. While R-squared indicates the proportion of variation in the dependent variable that can be explained by the independent variables, adjusted R-squared takes into account the number of predictors used in the model.

Adjusted R-squared penalizes the presence of unnecessary predictors in the model by adjusting the value of R-squared based on the number of predictors and the sample size. Specifically, it takes into account the number of predictors in the model and reduces the value of R-squared when additional predictors do not significantly improve the fit of the model.

As more predictors are added to a model, the value of R-squared may increase, but this improvement may be due to chance rather than a true relationship between the predictors and the dependent variable. Adjusted R-squared addresses this issue by adjusting for the number of predictors in the model and providing a more accurate measure of the true relationship between the predictors and the dependent variable.

In summary, adjusted R-squared penalizes the presence of unnecessary predictors in the model by adjusting the value of R-squared based on the number of predictors and the sample size. It helps to avoid overfitting and provides a more accurate measure of the goodness of fit of the model.

11. Differentiate between Ridge and Lasso Regression.

Ans. **Lasso** is a modification of linear regression, where the model is penalized for the sum of absolute values of the weights. Thus, the absolute values of weight will be (in general) reduced, and many will tend to be zeros. During training, the objective function become:

$$\frac{1}{2m}\sum_{i=1}^{m}(y - Xw)^2 + alpha\sum_{j=1}^{p}\left|w_j\right|$$

As you see, Lasso introduced a new hyperparameter, *alpha*, the coefficient to penalize weights.

Ridge takes a step further and penalizes the model for the sum of squared value of the weights. Thus, the weights not only tend to have smaller absolute values, but also really tend to penalize the extremes of the weights, resulting in a group of weights that are more evenly distributed. The objective function becomes:

$$\sum_{i=1}^{n}(y - Xw)^2 + alpha\sum_{j=1}^{p}w_j^{\,2}$$

Lasso regularization tends to give sparse weights (most zeros), because the l1 regularization cares equally about driving down big weights to small weights or driving small weights to zeros. If you have a lot of predictors (features), and you suspect that not all of them are that important, Lasso and ElasticNet may be a really good idea to start with.

Ridge tends to give small but well distributed weights, because the l2 regularization cares more about driving big weight to small weights, instead of driving small weights to zeros. When there are only a few predictors, and we are confident that all of them should be really relevant for predictions, try Ridge as a good regularized linear regression method.

12. What is VIF? What is the suitable value of a VIF for a feature to be included in a regression modelling?

Ans. VIF (Variance Inflation Factor) is a measure of how much the variance of the estimated regression coefficient is inflated by the presence of correlated predictors in a linear regression model. It is used to detect the presence of multicollinearity, which occurs when two or more predictors in a model are highly correlated with each other.

VIF measures the extent to which the variance of the estimated regression coefficient of a particular predictor is inflated due to its correlation with other predictors. It is calculated for each predictor in the model, and a VIF value of 1 indicates no correlation between the predictor and other predictors in the model. A VIF value greater than 1 indicates that the predictor is correlated with other predictors in the model.

A commonly used rule of thumb is that a VIF value greater than 5 or 10 indicates a high degree of multicollinearity and should be a cause for concern. In this case, it may be necessary to re-evaluate the inclusion of the correlated predictors in the model or consider combining them into a single predictor.

However, the appropriate threshold for VIF values can depend on the specific context and the research question. In some cases, a lower or higher threshold may be appropriate. Additionally, it is important to consider other factors, such as the research question, the data, and the purpose of the analysis, when determining the suitability of including a predictor in a regression model.

In summary, VIF is a measure of the extent to which the variance of the estimated regression coefficient is inflated due to the presence of correlated predictors in a linear regression model. A VIF value greater than 5 or 10 indicates a high degree of multicollinearity, although the appropriate threshold can depend on the specific context and research question.

13. Why do we need to scale the data before feeding it to the train model?

Ans. Scaling the data before feeding it to a machine learning model is important for several reasons:

1. Better performance: Some machine learning algorithms, such as the k-nearest neighbors' algorithm and gradient descent-based algorithms, are sensitive to the scale of the input data. If the data is not scaled, the algorithm may give too much weight to features with larger scales, which can result in poorer performance. Scaling the data can help prevent this issue and improve the performance of the algorithm.
2. Faster convergence: Scaling the data can also help algorithms that use optimization techniques, such as gradient descent, converge faster. This is because when the input features are on different scales, the optimization process may take longer to find the optimal solution.
3. Regularization: Some regularization techniques, such as L1 and L2 regularization, penalize the magnitude of the coefficients of the input features. Scaling the data can help prevent some features from dominating the regularization term and improve the performance of the model.
4. Interpretation: Scaling the data can also make it easier to interpret the coefficients of the input features. When the features are on different scales, it can be difficult to compare the magnitude of the coefficients and determine which features are more important in predicting the target variable.

In summary, scaling the data before feeding it to a machine learning model is important for improving the performance and convergence of the algorithm, preventing regularization issues, and making it easier to interpret the coefficients of the input features.

14. What are the different metrics which are used to check the goodness of fit in linear regression?
Ans. Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. There are several metrics used to check the goodness of fit in linear regression. Some of the commonly used metrics are:

R-squared (R2): It is the proportion of the variance in the dependent variable that is predictable from the independent variable(s). R-squared values range from 0 to 1, with higher values indicating better fit.

Mean Squared Error (MSE): It measures the average squared difference between the predicted and actual values of the dependent variable. Lower values of MSE indicate better fit.

Root Mean Squared Error (RMSE): It is the square root of the MSE and represents the average difference between the predicted and actual values in the original units of the dependent variable. Lower values of RMSE indicate better fit.

Mean Absolute Error (MAE): It measures the average absolute difference between the predicted and actual values of the dependent variable. Lower values of MAE indicate better fit.

Residual Standard Error (RSE): It is the standard deviation of the residuals, which are the differences between the predicted and actual values of the dependent variable. Lower values of RSE indicate better fit.

Akaike Information Criterion (AIC): It is a measure of the relative quality of a statistical model, based on the trade-off between the goodness of fit and the complexity of the model. Lower values of AIC indicate better fit.

Bayesian Information Criterion (BIC): It is similar to AIC but penalizes the model more for increasing the number of independent variables. Lower values of BIC indicate better fit.

These metrics provide different aspects of the goodness of fit in linear regression, and the selection of the appropriate metrics depends on the research question and the context of the problem.