1) Explain the concept of R-squared in linear regression models. How is it calculated, and what does it

represent?

Ans- R-squared (R2) is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable in a regression model.

2) Define adjusted R-squared and explain how it differs from the regular R-squared.

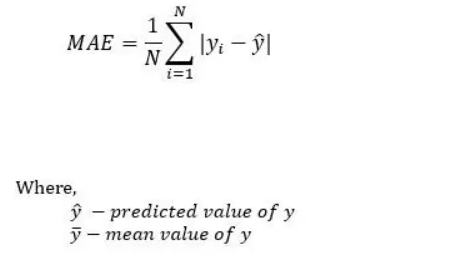
Ans- Adjusted R-squared is a modified version of R-squared that has been adjusted for the number of predictors in the model. The adjusted R-squared increases when the new term improves the model more than would be expected by chance. It decreases when a predictor improves the model by less than expected.

3) When is it more appropriate to use adjusted R-squared?

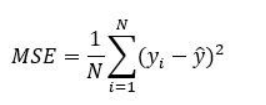
Ans- Clearly, it is better to use Adjusted R-squared when there are multiple variables in the regression model. This would allow us to compare models with differing numbers of independent variables.

4) What are RMSE, MSE, and MAE in the context of regression analysis? How are these metrics calculated, and what do they represent?

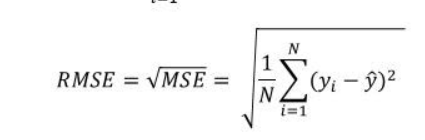
Ans- Usually the metrics used are the Mean Average Error (MAE), the Mean Squared Error (MSE) or the Root Mean Squared Error (RMSE). In short, MAE evaluates the absolute distance of the observations (the entries of the dataset) to the predictions on a regression, taking the average over all observations.



* Mean Squared Error represents the average of the squared difference between the original and predicted values in the data set. It measures the variance of the residuals.



* Root Mean Squared Error is the square root of Mean Squared error. It measures the standard deviation of residuals.



5) Discuss the advantages and disadvantages of using RMSE, MSE, and MAE as evaluation metrics in regression analysis.

* Ans- Mean Squared Error(MSE) and Root Mean Square Error penalizes the large prediction errors vi-a-vis Mean Absolute Error (MAE). However, RMSE is widely used than MSE to evaluate the performance of the regression model with other random models as it has the same units as the dependent variable (Y-axis).
* MSE is a differentiable function that makes it easy to perform mathematical operations in comparison to a non-differentiable function like MAE. Therefore, in many models, RMSE is used as a default metric for calculating Loss Function despite being harder to interpret than MAE.
* The lower value of MAE, MSE, and RMSE implies higher accuracy of a regression model. However, a higher value of R square is considered desirable.
* R Squared & Adjusted R Squared are used for explaining how well the independent variables in the linear regression model explains the variability in the dependent variable. R Squared value always increases with the addition of the independent variables which might lead to the addition of the redundant variables in our model. However, the adjusted R-squared solves this problem.
* Adjusted R squared takes into account the number of predictor variables, and it is used to determine the number of independent variables in our model. The value of Adjusted R squared decreases if the increase in the R square by the additional variable isn’t significant enough.

6) Explain the concept of Lasso regularization. How does it differ from Ridge regularization, and when is it more appropriate to use?

Ans- Similar to the lasso regression, ridge regression puts a similar constraint on the coefficients by introducing a penalty factor. However, while lasso regression takes the magnitude of the coefficients, ridge regression takes the square. Ridge regression is also referred to as L2 Regularization.

7) How do regularized linear models help to prevent overfitting in machine learning? Provide an example to illustrate.

Ans- Regularization in machine learning is the process of regularizing the parameters that constrain, regularizes, or shrinks the coefficient estimates towards zero. In other words, this technique discourages learning a more complex or flexible model, avoiding the risk of Overfitting.

8) Discuss the limitations of regularized linear models and explain why they may not always be the best choice for regression analysis.

Ans- This article discusses the problems that may occur while training a Linear model, and some methods to deal with them.

Non-Linearity of the response-predictor relationships.

Correlation of error terms.

A non-constant variance of the error term [Heteroscedasticity]

Collinearity.

Outliers and High Leverage Points.

9) You are comparing the performance of two regression models using different evaluation metrics.

Model A has an RMSE of 10, while Model B has an MAE of 8. Which model would you choose as the better

performer, and why? Are there any limitations to your choice of metric?

Ans- Both metrics have comparable behaviour in response to model bias and asymptote to the model bias as the bias increases. MAE is shown to be an unbiased estimator while RMSE is a biased estimator. MAE also has a lower sample variance compared with RMSE indicating MAE is the most robust choice.

10) You are comparing the performance of two regularized linear models using different types of

regularization. Model A uses Ridge regularization with a regularization parameter of 0.1, while Model B uses Lasso regularization with a regularization parameter of 0.5. Which model would you choose as the better performer, and why? Are there any trade-offs or limitations to your choice of regularization method?

Ans- regularization is used?

Regularization basically adds the penalty as model complexity increases. Regularization parameter (lambda) penalizes all the parameters except intercept so that model generalizes the data and won't overfit.