90-48=42

#1 X -10
#4 X -10
#7 -12
#8 X -10

Regression Analysis
Quiz-3

Likhit Garimella
lg836

Q.2)

(a) True
(b) True X
(c) True X
(d) True X
(e) True
(f) True
(g) True

-6

Q.3)

Here are three important characteristics to look for in a metric for cross-validation:

1. The metric should be relevant to the specific problem you are trying to solve. It should align with the goals and objectives of your analysis. For example, if you're working on a classification problem, F1-score would be relevant metrics to consider. If you're working on a regression problem, mean squared error (MSE) might be more appropriate.
2. The metric should be able to capture variations in the model's performance across different cross-validation folds. It should be able to differentiate between good and poor generalization. A metric that fails to distinguish between different model performances may not provide reliable insights. For example, if your model is sensitive to false positives, precision might be a more suitable metric than accuracy.
3. The metric should be more easily interpretable and understandable. This characteristic enables effective communication of the model's performance to other team members. Metrics like accuracy, precision, and recall are often more interpretable than complex metrics.

Q.4)

a third predictor (x3), the response variable (Y), a potential predictor (x3), the existing predictors (x1 and x2). Steps to create a partial residual plot:

OK

1. Fit the initial regression model using x1 and x2 as predictors and obtain the residuals, denoted as res.
2. Fit a new regression model by adding x3 as a predictor and obtain the residuals from this updated model, denoted as res_new.
3. Calculate the partial residuals by subtracting the residuals from the updated model (res_new) from the residuals of the initial model (res). These partial residuals represent the contribution of x3 to the overall residual variation.
4. Plot the partial residuals against x3. This plot will provide insights into the relationship between Y and x3 after accounting for the effects of x1 and x2.

Q.5)

Using scaled residuals (ei/s) instead of raw residuals (ei) in residual analysis offers several advantages:

1. Standardization: Scaling the residuals ensures that they have a standardized distribution with a mean of zero and a standard deviation of one. This standardization facilitates easier comparison and interpretation of the residuals, as they are now on a common scale.
2. Homoscedasticity assessment: Scaled residuals allow for a more reliable assessment of homoscedasticity (constant variance) assumptions. When using raw residuals, the spread of residuals may vary across the range of the predictor variables, making it difficult to detect heteroscedasticity. Scaling the residuals helps to overcome this issue by highlighting patterns or trends in the spread of residuals more effectively.
3. Outlier detection: Scaled residuals are particularly useful for identifying outliers or influential data points. Outliers often have a substantial impact on the model fit and can skew the analysis. By standardizing the residuals, outliers become more apparent as they deviate significantly from the expected range of values.
4. Diagnostic plots: When creating diagnostic plots, such as residual plots or normal probability plots, using scaled residuals ensures that the plots have consistent scales and easier interpretability. It allows for a better understanding of the distributional properties of the residuals and whether they conform to the assumptions of the statistical model.

In summary, using scaled residuals in residual analysis improves the reliability and interpretability of the results, facilitates the assessment of assumptions, aids in outlier detection, and enhances the visual diagnostics of the model.

Q.9)

two assumptions of the Simple Linear Model that appear to be violated are:

1. Linearity: violation of this assumption can be indicated by the presence of a non-linear pattern or curvature in the plot i.e. if the residuals exhibit a systematic pattern (e.g., a curved or U-shaped pattern) as the fitted values change, it suggests that the linear relationship assumption may not hold.
2. Homoscedasticity: in the residuals versus fitted values plot, violation of this assumption can be observed by the presence of a "fan-shaped" pattern or unequal spread of residuals around the fitted line.