SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore – 641 107
Accredited by NAAC-UGC with 'A' Grade
Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of Information Technology

Course Name – 23ADT202 Fundamental of Data science and Analytics

II Year / IV Semester

Unit 2 – Descriptive Analytics

Interpretation of r2

Introduction to Coefficient of Determination

The coefficient of determination (R^2) is a statistical metric that explains how well the independent variable(s) predict the dependent variable.

It ranges from 0 to 1, with higher values indicating a better model fit.

R^2 helps in evaluating the strength of the relationship between variables.

It is a crucial measure in regression analysis for model assessment

Introduction to Coefficient of Determination

R^2 represents the proportion of variance in the dependent variable that is explained by the independent variable(s).

A value of $R^2 = 1$ means perfect prediction with no error.

A value of $R^2 = 0$ indicates that the model does not explain any of the variance.

Values between 0 and 1 signify varying degrees of model fit.

Interpretation of $R^2 = 1$

When $R^2 = 1$, the regression model explains 100% of the variance in the dependent variable.

The model's predictions are perfectly accurate.

All data points lie directly on the regression line.

This indicates a flawless fit and no unexplained variance.

Interpretation of $R^2 = 0$

When $R^2 = 0$, the model explains none of the variance in the dependent variable.
The regression line does not provide any better predictions than simply using the mean of the dependent variable.
There is no linear relationship between the variables.
The model has no explanatory power.

Interpretation of R^2 Between 0 and 1

A value of R^2 between 0 and 1 shows the proportion of variance explained by the model.

Example: $R^2 = 0.8$ means the model explains 80% of the variance in the dependent variable.

The remaining 20% is unexplained, likely due to other factors.

Higher R^2 values indicate a stronger relationship and better fit.

Example of $R^2 = 0.7733$

In this example, the regression model explains 77.33% of the variance in the dependent variable.

The remaining 22.67% is unexplained, indicating areas where the model could be improved.

The model captures most, but not all, of the variance.

This suggests a moderately strong fit between the variables.

Visualizing the Coefficient of Determination

- Scatter plots with a regression line are an effective way to visualize how well a model fits the data.
- The closer the data points are to the regression line, the higher the R^2 value.
- A steep, straight line indicates a strong relationship, while scattered points indicate weak relationships.
- Visual tools help in interpreting the goodness of fit.

Applications of R^2

1. **Business:** Evaluate how well marketing efforts predict sales performance.
2. **Healthcare:** Measure the effectiveness of treatments in predicting recovery times.
3. **Education:** Assess the relationship between study hours and exam performance.

- R^2 can guide decision-making and improve predictions in various fields.

Exercise and Conclusion

1. **Exercise:** Calculate R^2 for X = [10, 20, 30, 40], Y = [15, 30, 25, 40], Y_pred = [14, 29, 26, 41].

2.Interpretation: Understand what the R^2 value indicates for the dataset.

- Calculating R^2 and interpreting its value helps assess model quality.
- A solid grasp of R^2 is essential for making informed decisions based on regression models.

THANK YOU