Metrics To Evaluate Regression

23 June 2021 11:13

1.R-Squared

About :

Percentage of the dependent variable variance that the model explains .How well model fits the data.

Pros

a. Simple and Intuitive. Works well for linear models.
b. Most commonly used Goodness of Fit Measure.

Cons

a. Not valid for Non Linear Models, Cannot use R-squared to compare performance of between a linear model to a nonlinear model.
b. Doesn't Tell about Model Bias.
c. Cannot be used to compare performance between linear and nonlinear models.
d. It has a Flaw that if we are adding redundant variables to the data, the value of R-squared does not decrease it either stays same or increase by any random correlation that variable has on target.
e. Based on a Research Article to look at the effect of using R-squared to assess the goodness-of-fit for models that are not linear.
   - R-squared is consistently high for both excellent and appalling models.
   - R-squared will not rise for better models all of the time.
   - If you use R-squared to pick the best model, it leads to the proper model only 28-43% of the time.

resources:

- https://bmcpharma.biomedcentral.com/articles/10.1186/1471-2210-10-6,
- https://statisticsbyjim.com/regression/r-squared-invalid-nonlinear-regression/

2.Adjusted R Squared

About :

Adjusted R Squared also indicates how well terms fit a curve or line, but adjusts for the number of terms in a model.
It will penalize you for adding independent variables.

Pros

a. Handles the Flaw in R squared and it changes only when a new features adds importance to the model.

Cons

a. Has Same Limitations as R squared.

3.Standard Error

About :

The standard error of the regression(S), also known as the standard error of the estimate , represents the average distance that the observed values fall from the regression line RMSE can be referred as standard error.

Pros

a. Simple and easy to calculate.
b. Can be used to compare all kinds of model to access their performance

Cons

a. Cannot determine Model Quality just from the Standard error.

4.Residual Plots(Analysis)

About :

These Plots are used to check assumptions of Linear regressions and also the quality of Fit visually by plotting Residuals vs Predicting Variables, Fitted vs residuals , predicted vs actual etc.

a. Like residuals are consistent with random error, If you start to see patterns residuals vs predicting variables plot you know something is wrong with your model.Just check that they are randomly scattered around zero for the entire range of fitted values. When the residuals centre on zero, they indicate that the model's predictions are correct on average rather than systematically too high or low.
b. Homoscedasticity - by plotting the fitted response values (as per the model) vs. the residuals.
c. Goodness of Fit - Predicted vs Actual.
d. Normality - Histogram and Q-Q plot of normalized residuals

Pros

a. Residual plots can expose a biased model far more effectively than the numeric output by displaying problematic patterns in the residuals. If your model is biased, you cannot trust the results.

Cons

a. Cannot generate a quality check number with some magnitude so it is Hard in Accessing and describing Quality between models

Resources :

- **https://www.statisticshowto.com/residual-plot/ ,**
- https://statisticsbyjim.com/regression/check-residual-plots-regression-analysis/

5.AIC/BIC(Probabilistic Model Selection)

About :

provides an analytical technique for scoring and choosing among candidate models. Here Model Selection is achieved by combining the complexity of the model with the performance of the model into a score, Here Penalty is Added when model is More Complex

Pros

a. Takes into account of **model performance and complexity** while another model selection technique of resampling(Train, Validation, and Test dataset) checks only model performance.
b. Doesn't need Test data , all the data can be used to train the model.
c. These Methods are particularly valuable for time series, because time series analysis' most valuable data is often the most recent, which is stuck in the validation and test sets.

Cons

AIC only measures the relative quality of models. This means that all models tested could still fit poorly. As a result, other measures are necessary to show that your model's results are of an acceptable absolute standard.

Resources :

- **https://towardsdatascience.com/introduction-to-aic-akaike-information-criterion-9c9ba1c96ced ,**
- https://machinelearningmastery.com/probabilistic-model-selection-measures/

6. Conclusion

Since Probabilistic methods evaluate best model in a group of models and are more useful when we don't have sufficient test d ata we can ignore measuring them for now. At this point of time to access model quality and compare performance between models we could use Standard Error and Residual plots R2 and Adjusted R2 can be chosen as secondary metric where we can give adjusted R2 priority when comparing similar kind(linear, Non-Linear) models.