
Assignment 8

1. What are some benefits of feature selection? How do you use the F-test to select the features?

Feature selection helps reduce the dimensionality of the dataset and reduce model complexity.

- **Reduces Overfitting:** Less redundant data means less opportunity to make decisions based on noise.
- **Improves Accuracy:** Less misleading data means modeling accuracy improves.
- **Reduces Training Time:** Less data means that algorithms train faster.

An F-statistic, or F-test, is a class of statistical tests that calculate the ratio between variances values, such as the variance from two different samples or the explained and unexplained variance by a statistical test, like ANOVA.

- ANOVA tests the relationship between categorical predictor vs continuous response.
- Check for equal variance between groups of categorical feature wrt continuous response.
- If there is equal variance between groups, it means this feature has no impact on response and it can be ignored in model training.

2. Can we use PCA for feature selection? If yes, then why?

The only way PCA is a valid method of feature selection is if the most important variables are the ones that happen to have the most variation in them.

PCA helps in data transformation and prevents data loss, which is great for dimension reduction and could result in better regression models.

3. What's the difference between forwarding Feature Selection and Backward Feature Selection?

Forward Selection: Forward selection is an iterative method in which we start with having no feature in the model. In each iteration, we keep adding the feature which best improves our model till an addition of a new variable does not improve the performance of the model.

Backward selection starts with all features contained in the dataset. It then runs a model and calculates a p-value associated with the t-test or F-test of the model for each feature. The feature with the **largest insignificant p-value** will then be removed from the model, and the process starts again.

4. How do you transform a skewed distribution into a Normal Distribution? Name some techniques?

Taking the **square root** and the **logarithm** of the observation in order to make the distribution normal belongs to a class of transforms called power transforms.

The Box-Cox method is a data transform method that is able to perform a range of power transforms, including the log and the square root.

Some of the transformation techniques are:

- Log Transformation.
- Square root Transformation.
- Reciprocal Transformation.
- Exponential Transformation.
- Box-Cox Transformation.

5. How to perform Feature Engineering on Unknown features?

- Handle missing values.
- Outliers.
- Scaling.
- Encoding.
- Balance & imbalance dataset.
- Transformation.
- Feature Selection.
- Dimensionality reduction

