**Assignment 8**

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

Ans:

By using feature selection only certain features which are most important predictors of the target variable are used. This way eliminating some irrelevant, unimportant features while building model.

F Test is a statistical test used to compare between models and check if the difference is significant between the models.

F-Test does a hypothesis testing model X and Y where X is a model created by just a constant and Y is the model created by a constant and a feature.

The least square errors in both the models are compared and checks if the difference in errors between model X and Y are significant or introduced by chance.

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

Ans:

PCA is a dimensionality reduction technique. PCA constructs relevant features/variables through linear /non-linear combinations of Original features.

PCA technique is particularly useful in processing data where multi-collinearity exists between the features/variables.

PCA can be used when the dimensions of the input features are high (e.g. a lot of variables).

PCA can be also used for denoising and data compression.

So, yes PCA can be used as a feature selection technique.

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

Ans:

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 Elimination: In backward elimination, we start with all the features and removes the least significant feature at each iteration which improves the performance of the model. We repeat this until no improvement is observed on removal of features.

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

Ans: Following are the techniques for Transforming skewed to Normal Distribution,

Log Transformation:

Numerical variables may have high skewed and non-normal distribution (Gaussian Distribution) caused by outliers, highly exponential distributions, etc. Therefore we go for data transformation.

In Log transformation each variable of x will be replaced by log(x) with base 10, base 2, or natural log.

Square-Root Transformation:

This transformation will give a moderate effect on distribution. The main advantage of square root transformation is, it can be applied to zero values.

Here the x will replace by the square root(x). It is weaker than the Log Transformation.

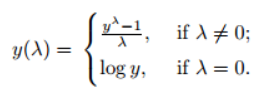
Reciprocal Transformation:

In this transformation, x will replace by the inverse of x (1/x).

The reciprocal transformation will give little effect on the shape of the distribution. This transformation can be only used for non-zero values.

Box-Cox Transformation:

Box-cox transformation works pretty well for many data natures. The below image is the mathematical formula for Box-cox transformation.



All the values of lambda vary from -5 to 5 are considered and the best value for the data is selected. The “Best” value is one that results in the best skewness of the distribution. Log transformation will take place when we have lambda is zero.

Exponential Transformation:

The exponential value of the Price variable will be taken.

5. How to perform Feature Engineering on Unknown features?

Ans:

Various Automated Feature engineering techniques can be used which required minimal domain knowledge. Few Automated Feature engineering techniques,

* Principal component analysis (PCA) and independent component analysis (ICA) map existing data to another feature space
* Deep feature synthesis (DFS) allows for transfer of intermediate learnings from middle layers in the neural networks.