**Feature selection by using Mutual Information**

MI Estimate mutual information for a discrete target variable.

Information gain can also be used for feature selection, by evaluating the gain of each variable in the context of the target variable. In this slightly different usage, the calculation is referred to as mutual information between the two random variables.

Mutual information (MI) between two random variables is a non-negative value, which measures the dependency between the variables. It is equal to zero if and only if two random variables are independent, and higher values mean higher dependency.

The function relies on nonparametric methods based on entropy estimation from k-nearest neighbors distances.

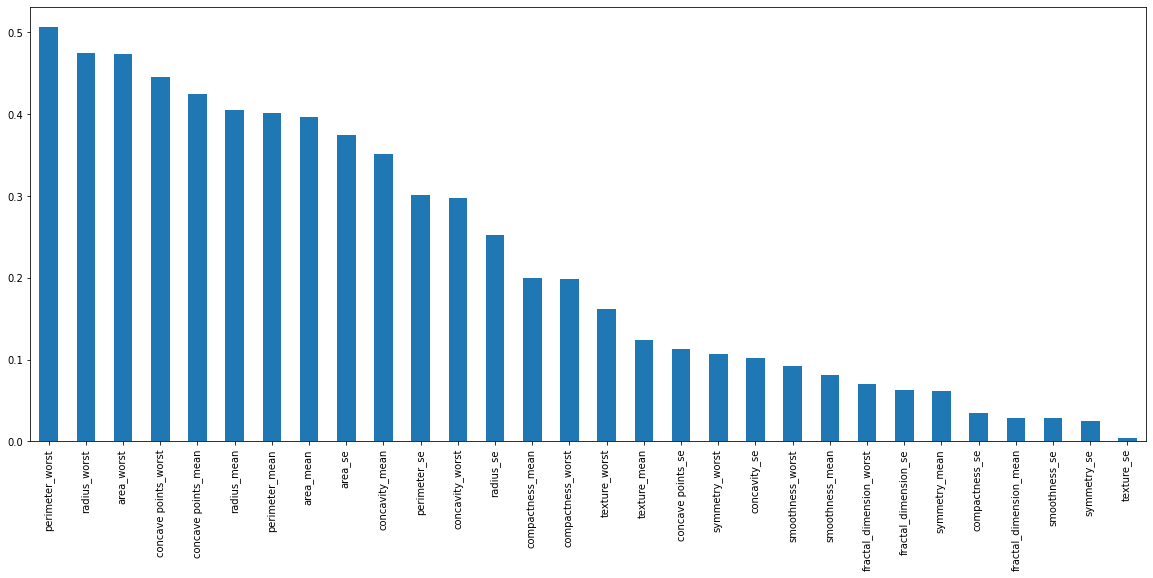
Inshort

A quantity called mutual information measures the amount of information one can obtain from one random variable given another.

The mutual information between two random variables X and Y can be stated formally as follows:

I(X ; Y) = H(X) – H(X | Y)

Where I(X ; Y) is the mutual information for X and Y, H(X) is the entropy for X and H(X | Y) is the conditional entropy for X given Y. The result has the units of bits.



**# By this analysis most appropriate features are:**

['radius\_mean', 'perimeter\_mean',

'area\_mean', 'concave points\_mean',

'radius\_worst', 'perimeter\_worst',

'area\_worst', 'concave points\_worst']

**# feature selection using Anova test**

Input\_Features Score P\_Value

27 concave points\_worst 633.821528 0.0

22 perimeter\_worst 620.150300 0.0

20 radius\_worst 593.498524 0.0

7 concave points\_mean 566.943032 0.0

2 perimeter\_mean 514.424407 0.0

0 radius\_mean 481.179225 0.0

23 area\_worst 448.161823 0.0

3 area\_mean 424.524809 0.0

Conclusions : Using Anova test and feature scaling using Mutual information we got the same result , same set of common important features . so we can concider these 8 features as our prime features and used for model training.