

CHI-SQUARED

FOR FEATURE SELECTION

To use χ^2 for feature selection, we calculate χ^2 between each feature and the target, and select the desired number of features with the best χ^2 scores.

The intuition is that if a feature is independent to the target it is uninformative for classifying observations.

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

of observations in class i

of expected observations in class i if there was no relationship between the feature and target.