<http://blog.datadive.net/selecting-good-features-part-i-univariate-selection/>

**Feature Selection – Univariate selection**

Two reasons why feature selection is used;

1. Reducing the no of features, to reduce overfitting and improving generalization of models.
2. Gain better understanding of features and their relationship to the response variables.

Linear model and regularization:

When there are multiple (linearly) correlated features (as is the case with very many real life datasets), the model becomes unstable, meaning that small changes in the data can cause large changes in the model (i.e. coefficient values), making model interpretation very difficult (so called [multicollinearity](http://en.wikipedia.org/wiki/Multicollinearity) problem).

Regularization models;

Regularization is a method for adding additional constraints or penalty to a model, with the goal of preventing overfitting and improving generalization.

There are 2 types of regularizations;

1. L1 (Lasso)
2. L2 (Ridge)

L1 Regularization:

It ensures that the model doesn’t overfit the data, essentially you are desensitizing your model to the training data. It can also help solve unsolvable equations.

It adds a penalty to loss function. Since each non-zero coefficient adds to the penalty, it forces weak features to have zero as coefficients. Thus L1 regularization produces sparse solutions, inherently performing feature selection.

The below video explains the purpose of Ridge regularization.

<https://www.youtube.com/watch?v=Q81RR3yKn30>