# The Curse of Dimensionality

If the number of features used in the data increases so**,**

-Number of samples should increase

-The possibility of overfitting increases

**How to reduce number of dimension of dataset?**

-Applying features selector "Variance Threshold"

**How to know correlation between features ?**

correlation not causation

using corr() method

then using from numby triu() to create matrix for true value with same dimension

r = -1 -> perfect negative correlation

r=0 -> No correlation

r=1 -> perfect positive correlation

so , drop on of two features that has the information

**Feature Selection Algorithim**

-Recursive feature elimination : drop features with less coef

Random forest algorithm manage to calculate feature importance values

random forest is combination of decision tree

we can use combination of models for feature selection

using feature_importances_ attribute

in linear regression

-we using lasso algo to avoid overfitting

lassoCV regressor -> choose optimal value for alpha

-RandomForestRegressor(RFE)

-GradientBoostingRegressor