**Salary Dataset - Simple linear regression (from Kaggle)**

Simple Linear Regression Dataset, used in Machine Learning A - Z

About Dataset

Dataset Description

Salary Dataset in CSV for Simple linear regression. It has also been used in Machine Learning A to Z course of my series.

Columns

1. #, 2. YearsExperience, 3. Salary

**Linear Regression Dataset (from Kaggle)**

About Dataset

Context

This is probably the dumbest dataset on Kaggle. The whole point is, however, to provide a common dataset for linear regression. Although such a dataset can easily be generated in Excel with random numbers, results would not be comparable.

Content

The training dataset is a CSV file with 700 data pairs (*x*, *y*). The *x*-values are numbers between 0 and 100. The corresponding *y*-values have been generated using the Excel function *NORMINV*(*RAND*(), *x*, 3). Consequently, the best estimate for *y* should be *x*.

The test dataset is a CSV file with 300 data pairs.

**Red Wine Quality Dataset （UCI Dataset）**

## About Dataset

### Context

The two datasets are related to red and white variants of the Portuguese "Vinho Verde" wine. For more details, consult the reference [Cortez et al., 2009]. Due to privacy and logistic issues, only physicochemical (inputs) and sensory (the output) variables are available (e.g. there is no data about grape types, wine brand, wine selling price, etc.).

These datasets can be viewed as classification or regression tasks. The classes are ordered and not balanced (e.g. there are much more normal wines than excellent or poor ones).

Content

Input variables (based on physicochemical tests):  
1 - fixed acidity, 2 - volatile acidity, 3 - citric acid, 4 - residual sugar, 5 – chlorides, 6 - free sulfur dioxide, 7 - total sulfur dioxide, 8 – density, 9 – pH, 10 – sulphates, 11 – alcohol  
  
Output variable (based on sensory data): 12 - quality (score from 3 to 8)