Machine Learning Course

Data prepossessing is written the notebook

This the Data Pre processing Template:

#Basic packages for DataScience

import numpy as np

import matplotlib.pyplot as plt

import pandas as pd

#import dataset as DataFrames and split them into independent and dependent variables

dataset = pd.read_csv('Salary_Data.csv')

years_exp = dataset.iloc[:, :-1]

salary_set = dataset.iloc[:,-1:]

#Split Data into Training set(used for training the ML model) and Test Set(Used to compare if the model is working as expected)

from sklearn.model_selection_selection import train_test_split

years_exp_train, years_exp_test, salary_set_train, salary_set_test = train_test_split(years_exp,salary_set,test_size=1/3,random_state = 0)

feature scaling module is need to normalize certain values so that they do not diminish the "Weight" of other values

"""

#Not always necessary

from sklearn.preprocessing import StandarScaler

sc_Independent = StandardScaler()

Independent_train = sc_Independent.fit_transform(Independent_train)

Independent_test = sc_Independent.transform(Independent_test)

sc_Dependent = StandardScaler()

Dependent_train = sc_Dependent.fit_transform(Dependent_train)

"""

Linear Regression

$$Y = b_0 + b_{1^*}x_1$$

This regression is used to map the minimum total distance between all the observations

$$\min\left(\sum (y_i - \tilde{y_i})^2\right)$$

The following code is used to get the linear regression model on your data set