

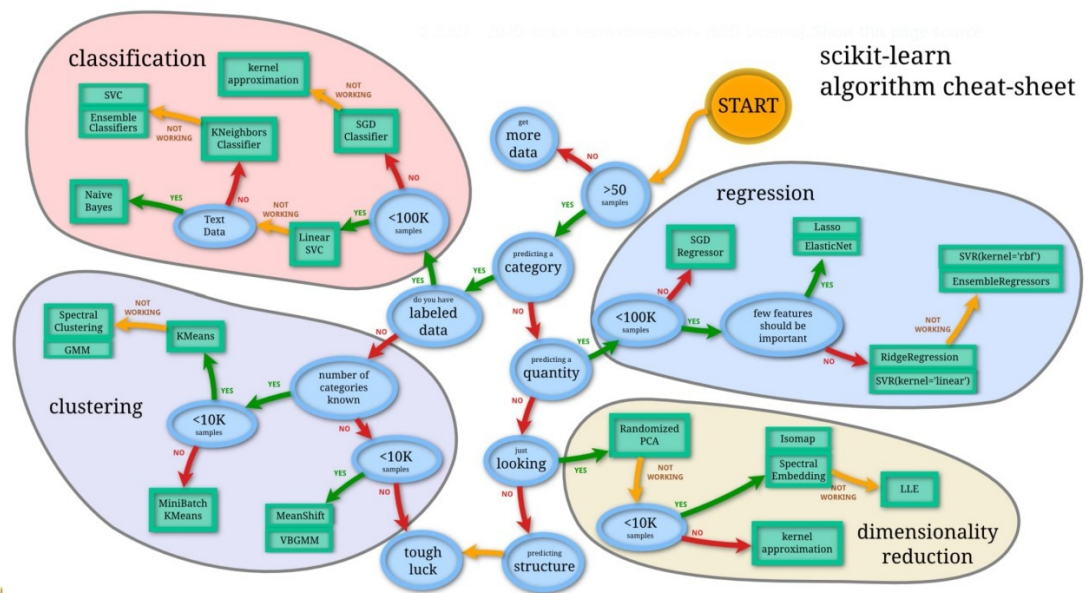
piscine-machine-learning-day03

November 23, 2023

0.1 REGRESSION

Scikit-learn is an open-source machine learning library. It features various classification, regression and clustering algorithms such as support vector machines (SVM), random forests, k-means clustering.

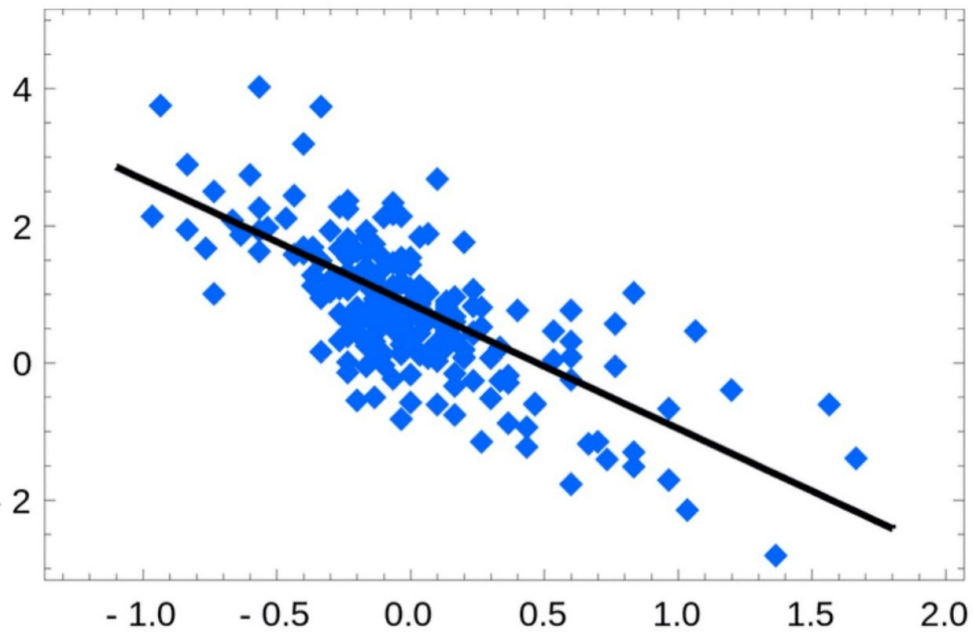
This library is maintained by INRA, a French Research Lab. Scikit-learn v1 was release in 2010 and as of today is considered to be one of the best framework to getting started with Machine Learning.



Sklearn-cheat

0.1.1 LINEAR REGRESSION

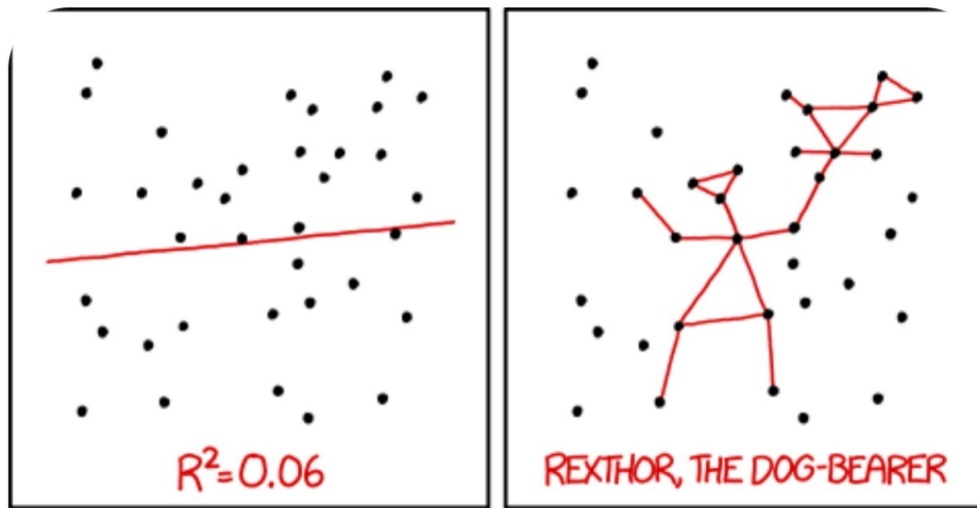
Regression is the task of predicting a continuous value. One typical use is predicting the price of the stock market or the price a house. In this example we want to find the line that cross most points form the data Regression :



Task01

You must try and plot all combinaison to determine which features are best to predicting the output. This is called feature selection

- 1 - Create a linear regression model using LinearRegression from Scikit-learn
- 2 - Plot the test set and the prediction of your model to compare them
- 3 - Display the performance indicators of your model : R^2

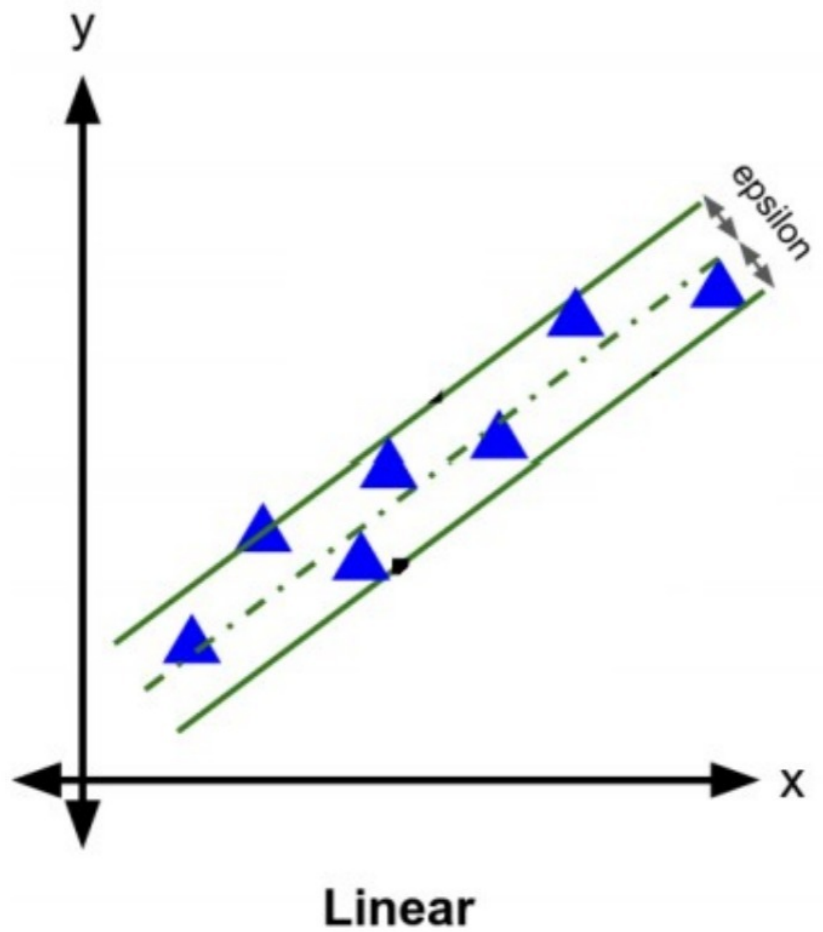


I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER
TO GUESS THE DIRECTION OF THE CORRELATION FROM THE
SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

R^2

0.1.2 SUPPORT VECTOR REGRESSOR

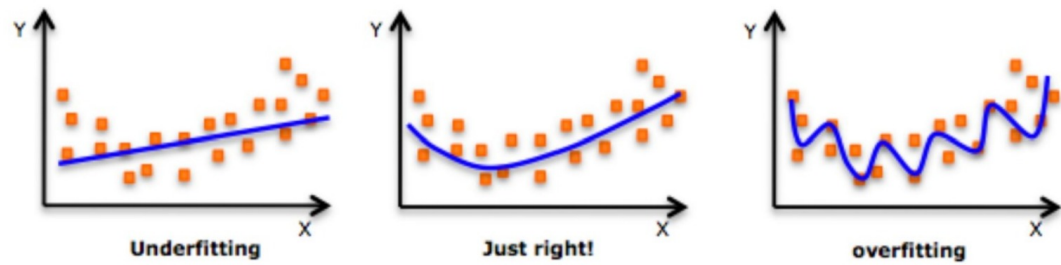
SVR is a regressor that is used to predict continuous ordered variables.



SUPPORT VECTOR REGRESSOR

Task02

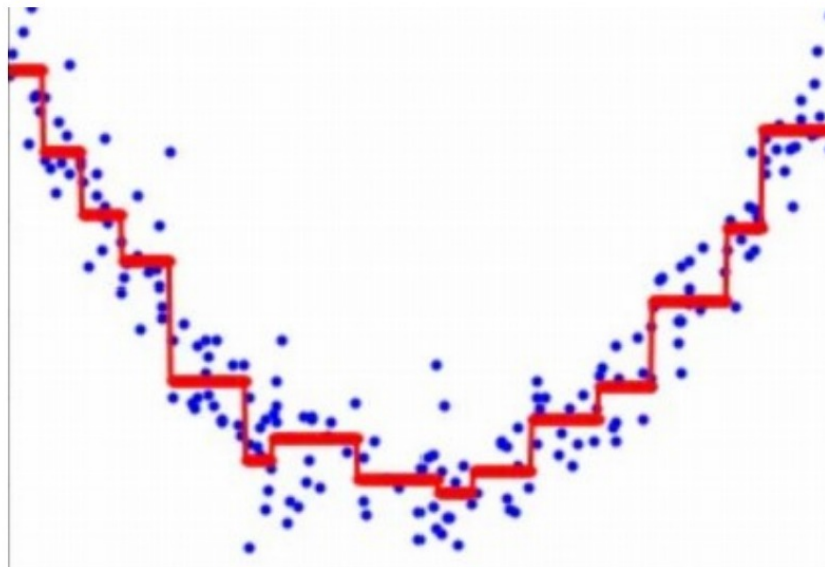
- 0 - When using SVM you must use Standard Scaldard before
- 1 - Create a LinearSVR model from Scikit-learn
- 2 - Plot the test set and the prediction of your model to compare them
- 3 - Display the performance indicators of your model : R^2
- 4 - Try differents values of the hyper-parameter C to tune your model.



Performance :

0.1.3 DECISION TREES REGRESSOR

Decision trees can also be used for regression tasks.



Performance :

Task03

- 1 - Create a DecisionTreeRegressor model from Scikit-learn
- 2 - Plot the test set and the prediction of your model to compare them
- 3 - Display the performance indicators of your model : R^2
- 4 - Try different values of the hyper-parameter max_depth to get better score. Which value is best ?

[]: