

Dominik Wróbel

Inżynieria oprogramowania i systemów

Informatyka, II stopień, 2018/19

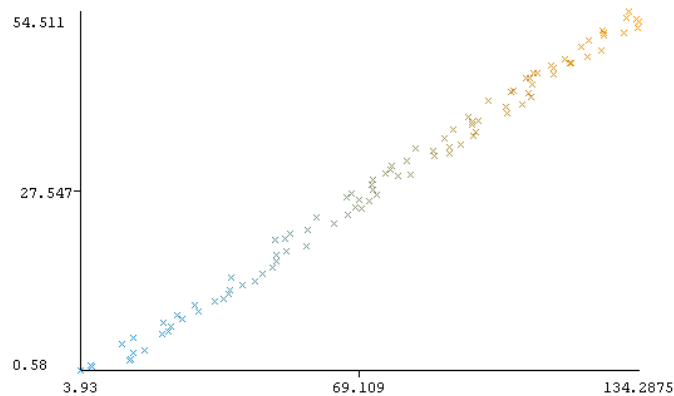
Metody eksploracji danych

Laboratorium 1

Weka + Python+ regresja

1.1

W oprogramowaniu Weka załadowano plik xy-001.arff z danymi i wyświetlono dane.

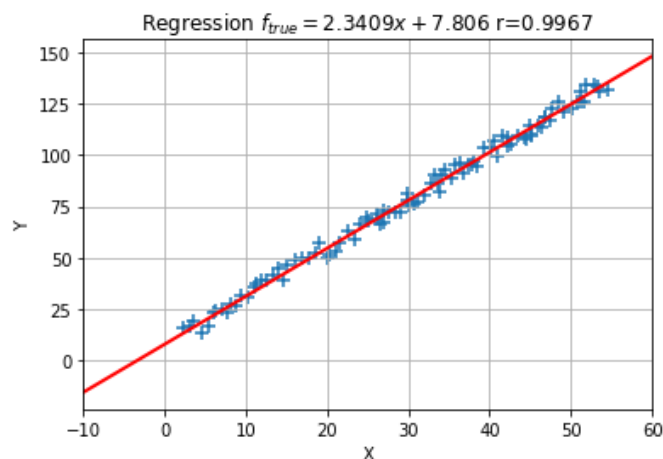


Dla przedstawionych danych wykonano regresję liniową i otrzymano równanie prostej oraz współczynnik korelacji równe

$$Y = 2.3409 * X + 7.806, r = 0.9967$$

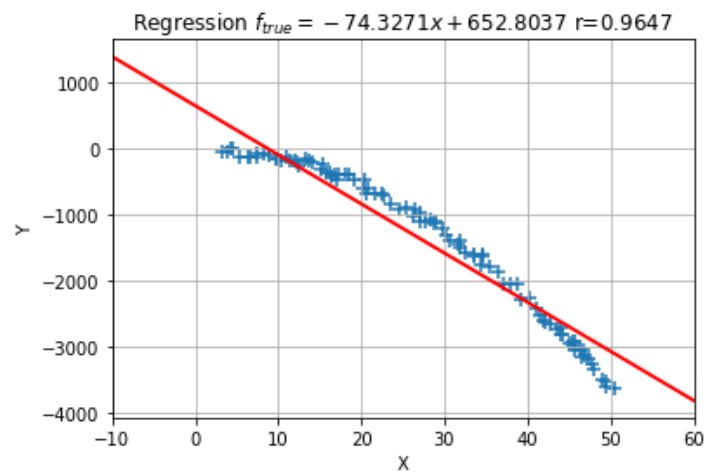
1.2

W kolejnych zadaniach wykonano te same czynności. Uzyskane w wyniku regresji proste wraz z danymi przedstawiono na wykresach z wykorzystaniem skryptu w języku Python. Plik xy-001.arff.



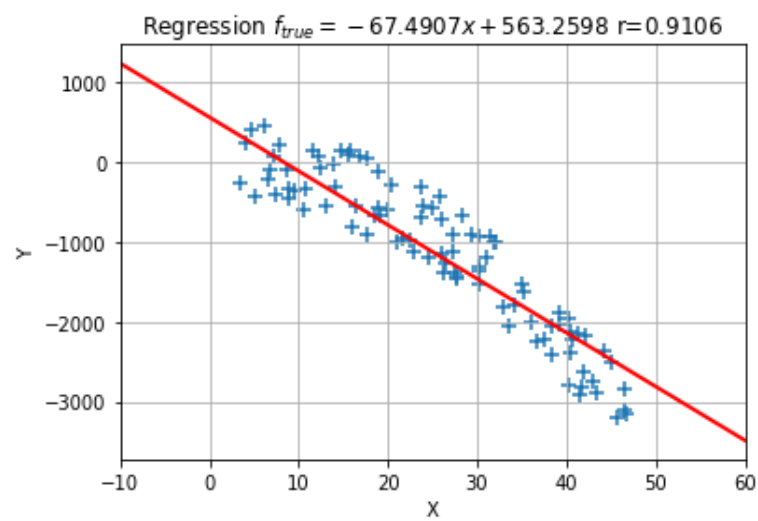
1.3

Plik xy-002.arff. $Y = -74.3271 * X + 652.8037$, $r = 0.9647$



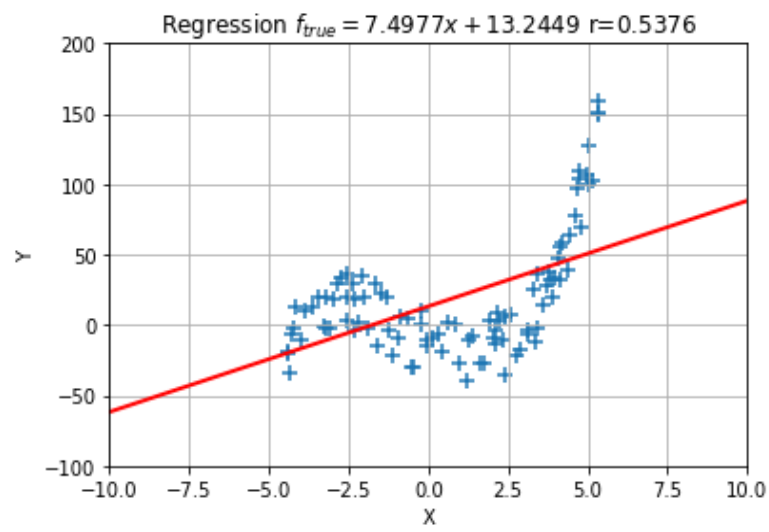
1.4

Plik xy-003.arff. $Y = -67.4907 * X + 563.2598$, $r = 0.9106$



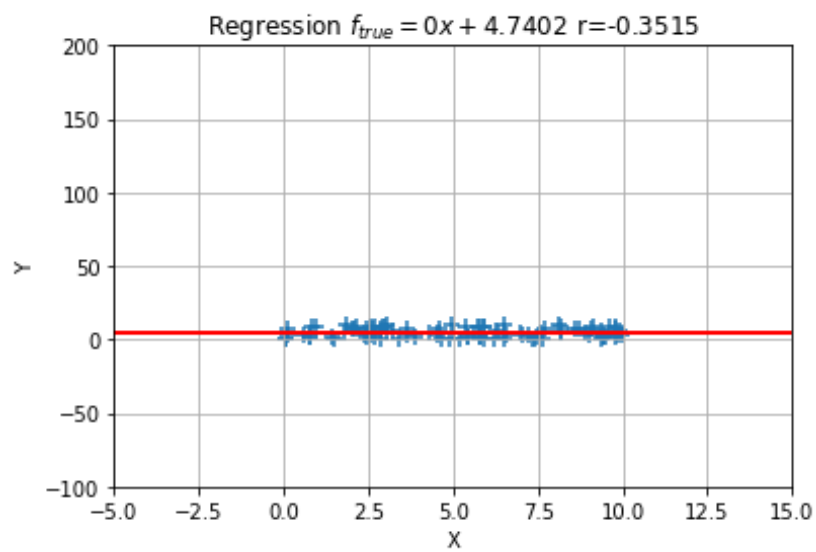
1.5

Plik xy-005.arff. $Y = 7.4977 * X + 13.2449$, $r = 0.5376$



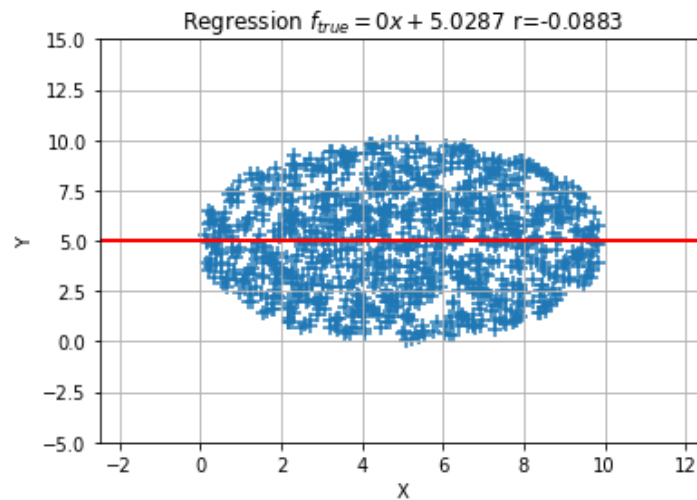
1.6

Plik xy-006.arff. $Y = 0 * X + 4.7402$, $r = -0.3515$



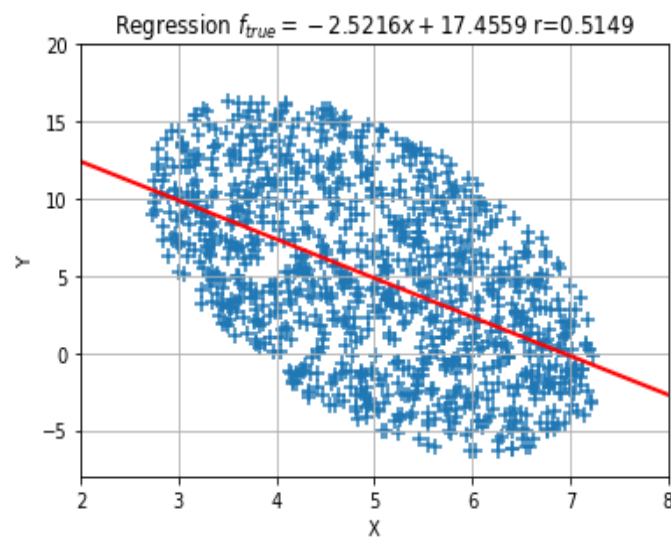
1.7

Plik xy-007.arff. $Y = 0 \cdot X + 5.0287$, $r = -0.0883$



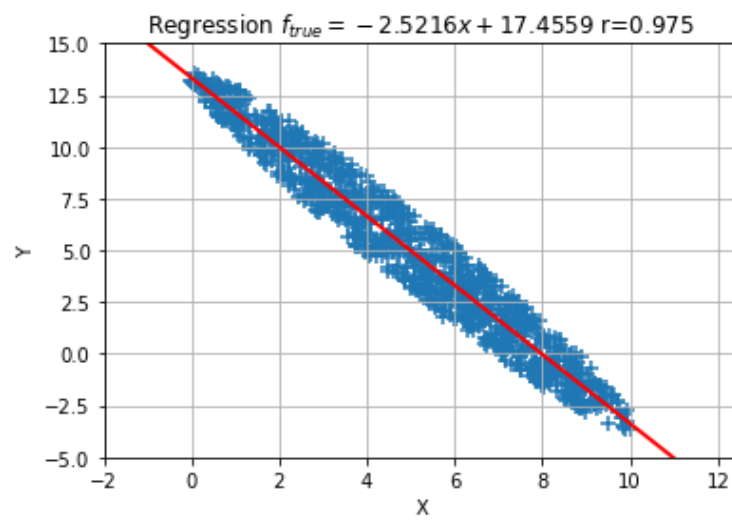
1.8

Plik xy-008.arff. $Y = -2.5216 \cdot X + 17.4559$, $r = 0.5149$



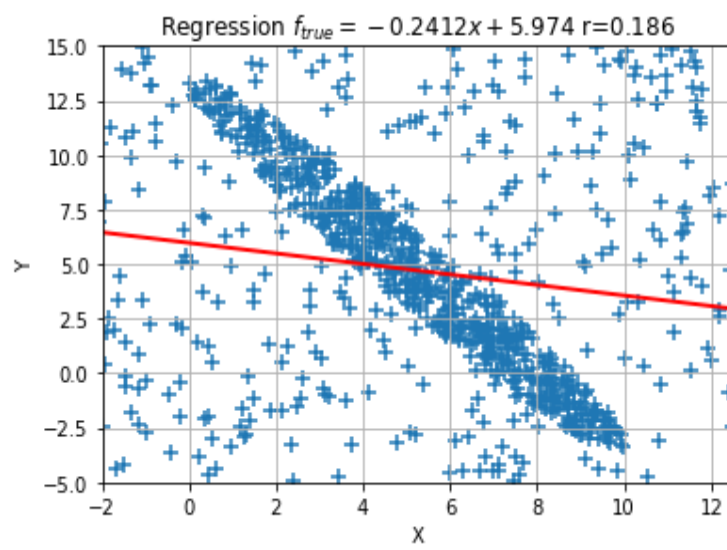
Komentarz: W tym przypadku prosta nie przechodzi przez oś symetrii elipsy. Wynika to ze specyfiki działania regresji liniowej. Algorytm oblicza odległość od prostej wyznaczaną względem osi pionowej, a nie prostopadłą do prostej.

Plik xy-009.arff. $Y = -1.6706 * X + 13.3334$, $r = 0.975$



1.9

Plik xy-010.arff. $Y = -0.2412 * X + 5.974$, $r = 0.186$

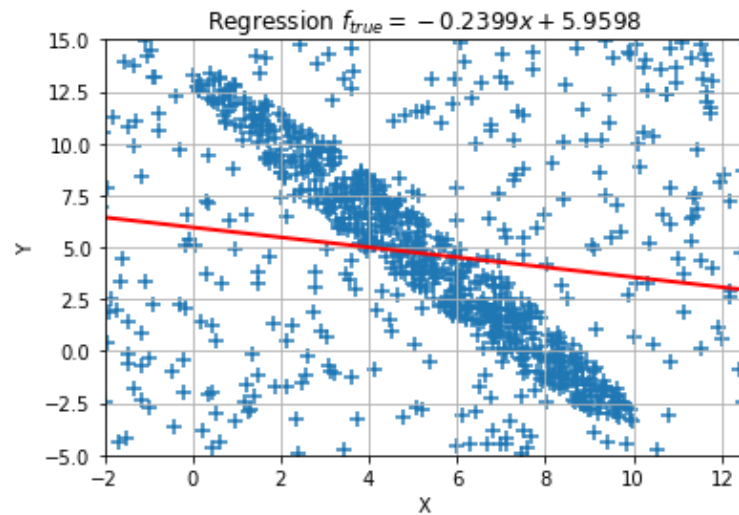


1.10

Plik xy-010.arff. Realizacja przy pomocy Pythona.

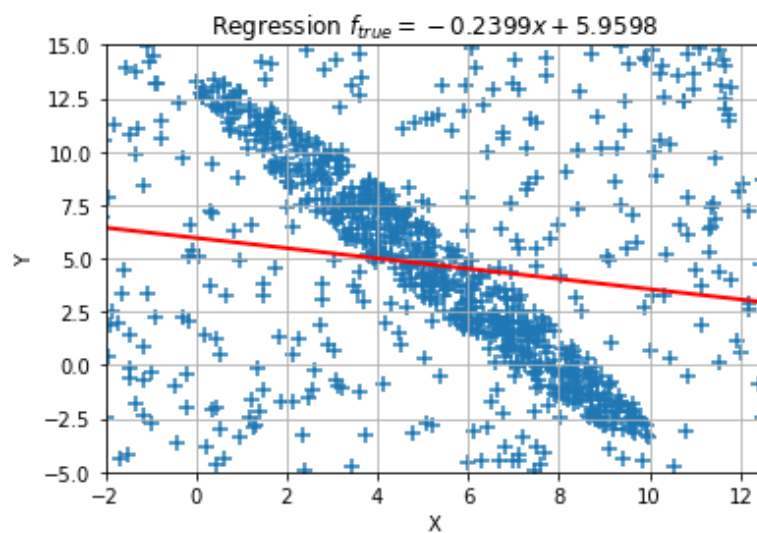
- Metoda najmniejszych kwadratów

$$Y = -0.2399 * X + 5.9598$$



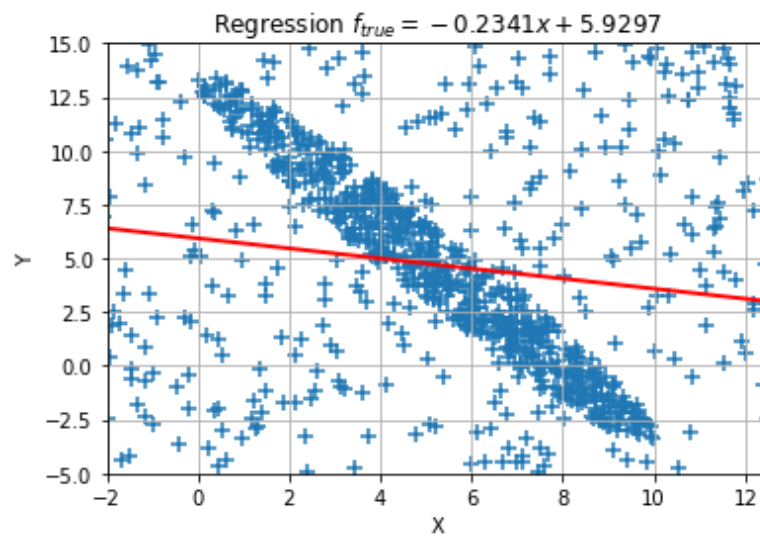
- Regresja grzbietowa (alpha= .5)

$$Y = -0.2399 * X + 5.9597$$



- Lasso

$$Y = -0.2341 * X + 5.9297$$



Wyniki uzyskane przy pomocy Pythona nie różnią się w dużym stopniu od tych uzyskanych przy pomocy oprogramowania Weka.