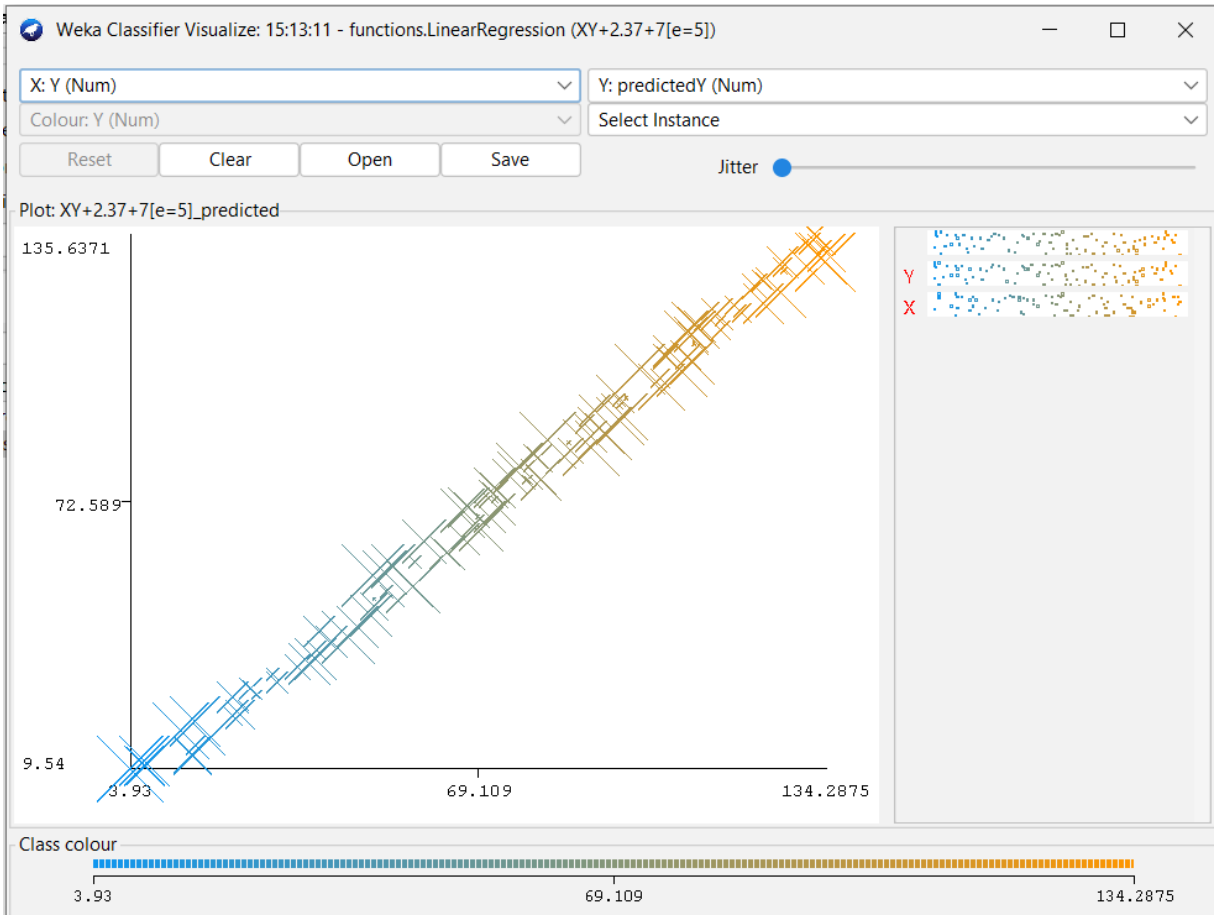
SPRAWOZDANIE – LABORATORIUM 1

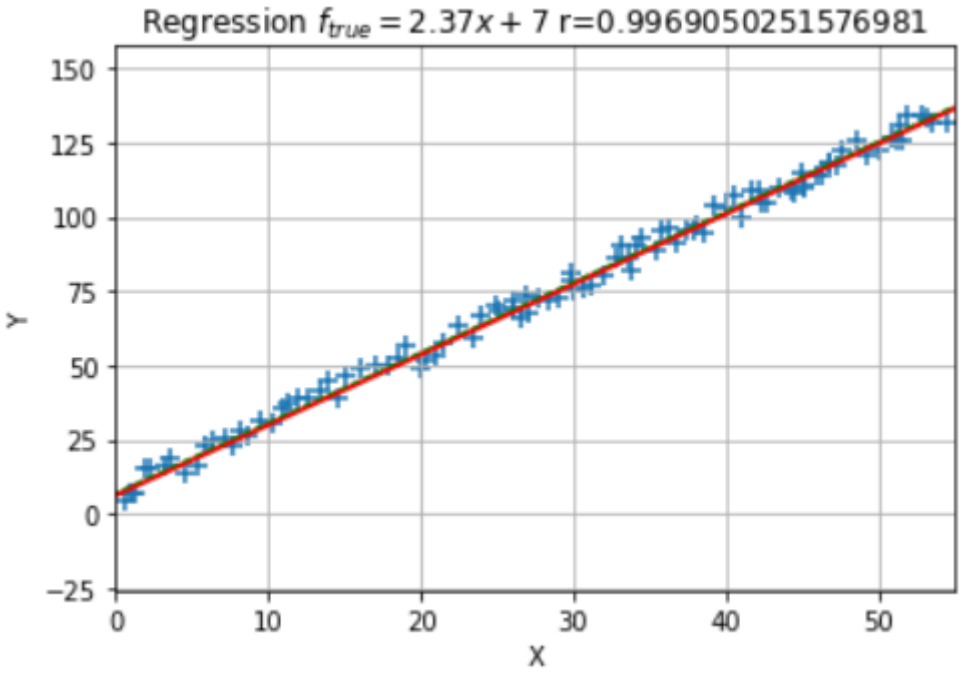
Karolina Kotłowska, 12 marca 2023

1.1



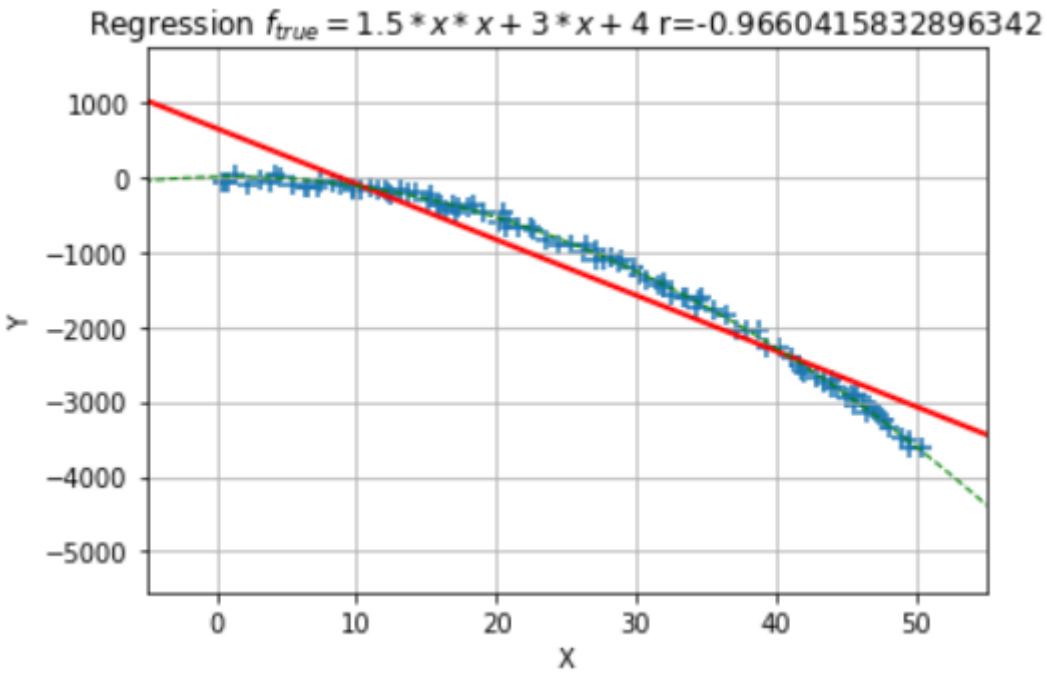
1.2

```
=== Run information ===

Scheme:       weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4
Relation:     XY+2.37+7[e=5]
Instances:    100
Attributes:   2
              X
              Y
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===


Linear Regression Model

Y =

      2.3409 * X +
      7.806

Time taken to build model: 0.05 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.9967
Mean absolute error                      2.6927
Root mean squared error                  3.0189
Relative absolute error                  8.3743 %
Root relative squared error              8.0309 %
Total Number of Instances              100
```



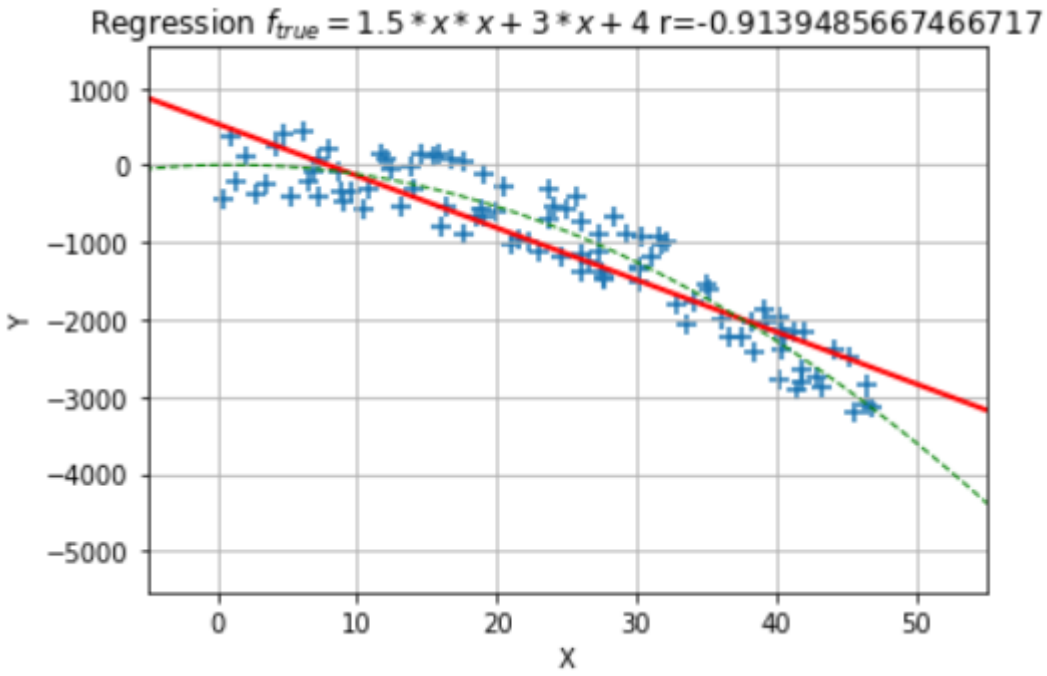
1.3

```
=== Run information ===

Scheme:       weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4
Relation:     XY-1.5X2+3X+4+7[e=100]
Instances:    100
Attributes:   2
              X
              Y
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===


Linear Regression Model

Y =

    -74.3271 * X +
    652.8037

Time taken to build model: 0 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.9647
Mean absolute error                    263.4154
Root mean squared error                302.3797
Relative absolute error                 25.8906 %
Root relative squared error             26.07   %
Total Number of Instances              100
```



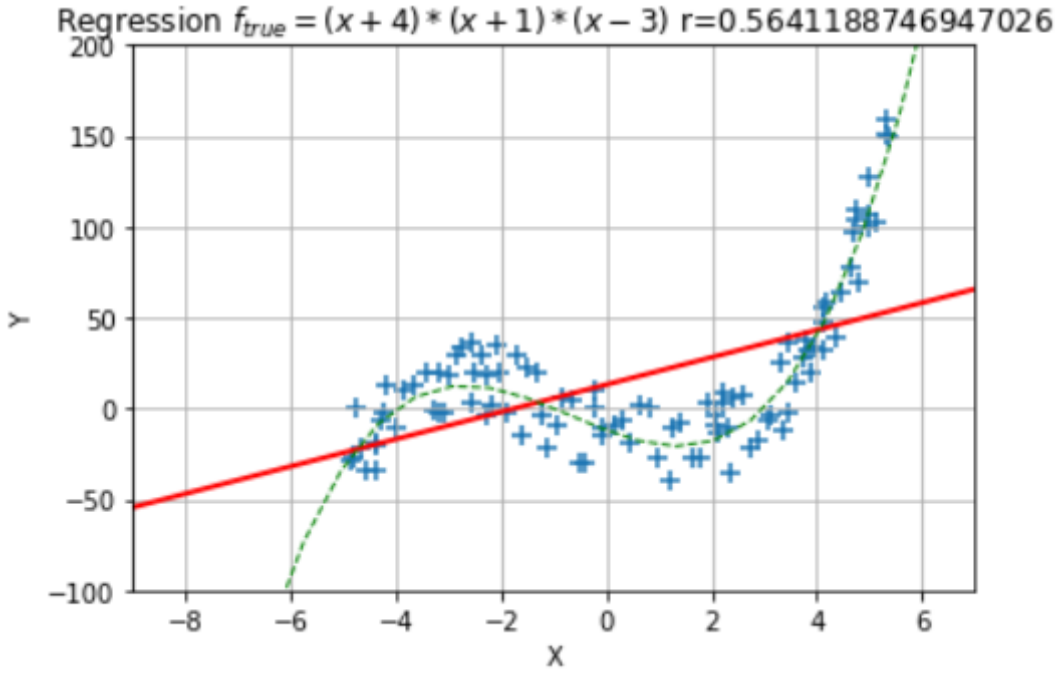
1.4

```
=== Run information ===

Scheme:       weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4
Relation:     XY-1.5X2+3X+4+7[e=500]
Instances:    100
Attributes:   2
              X
              Y
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===


Linear Regression Model

Y =

    -67.4907 * X +
    563.2598

Time taken to build model: 0 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.9106
Mean absolute error                    325.6369
Root mean squared error                401.8623
Relative absolute error                 39.4346 %
Root relative squared error             40.96   %
Total Number of Instances              100
```



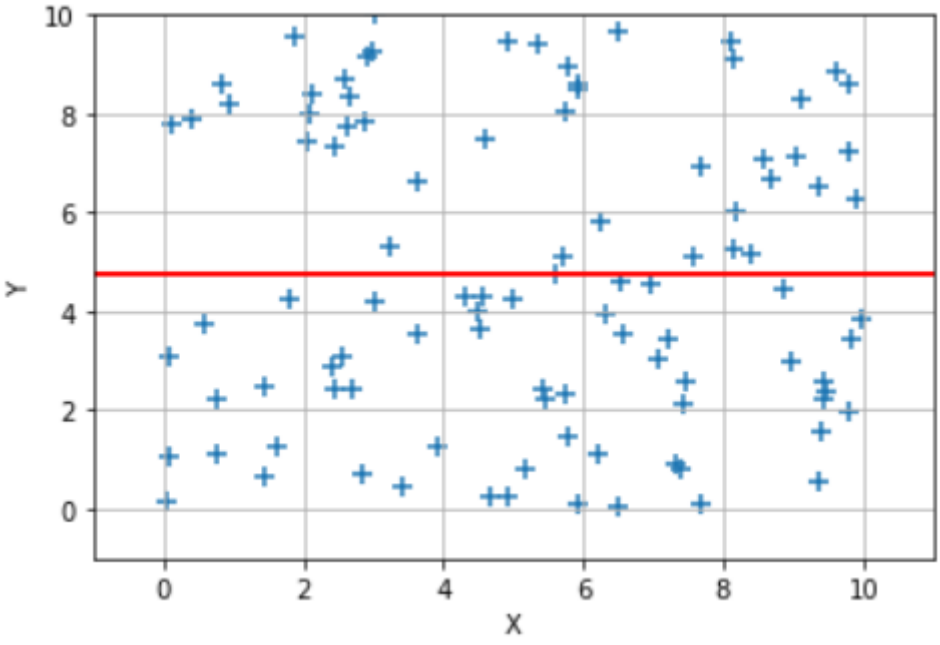
1.5

```
=== Run information ===

Scheme:       weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4
Relation:     XY-(x+4)(x+1)(x-3)+[e=25]
Instances:    100
Attributes:   2
              X
              Y
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===


Linear Regression Model

Y =

      7.4977 * X +
     13.2449

Time taken to build model: 0 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.5376
Mean absolute error                     28.4427
Root mean squared error                 35.9572
Relative absolute error                 89.5158 %
Root relative squared error             83.311  %
Total Number of Instances              100
```



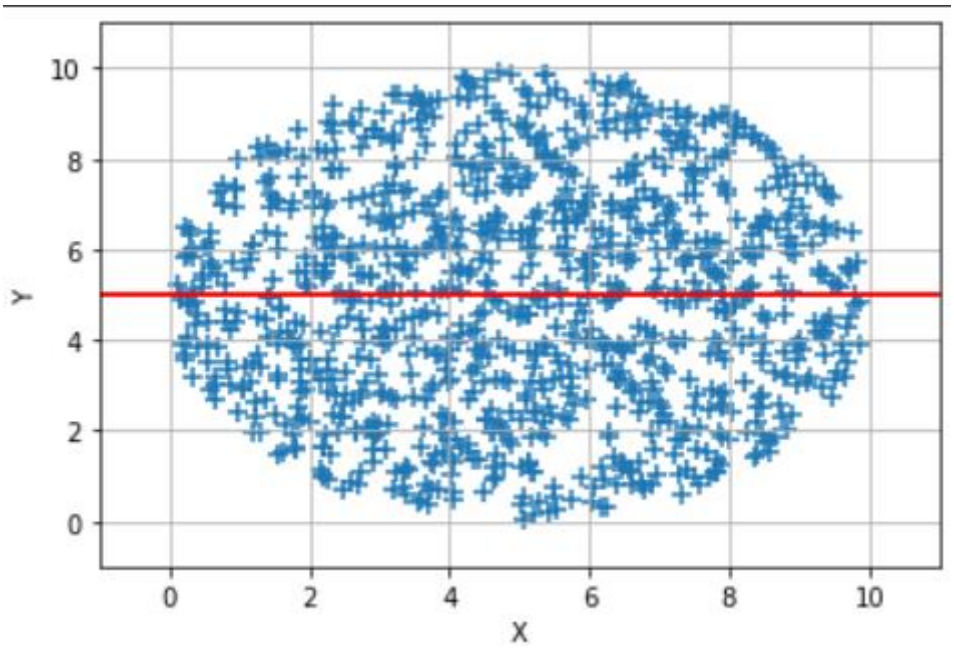
1.6

```
=== Run information ===

Scheme:       weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4
Relation:     XY-(5-5-5-box)
Instances:    100
Attributes:   2
              X
              Y
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===


Linear Regression Model

Y =

 +
      4.7402

Time taken to build model: 0 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                 -0.3515
Mean absolute error                      2.6968
Root mean squared error                  3.0565
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              100
```



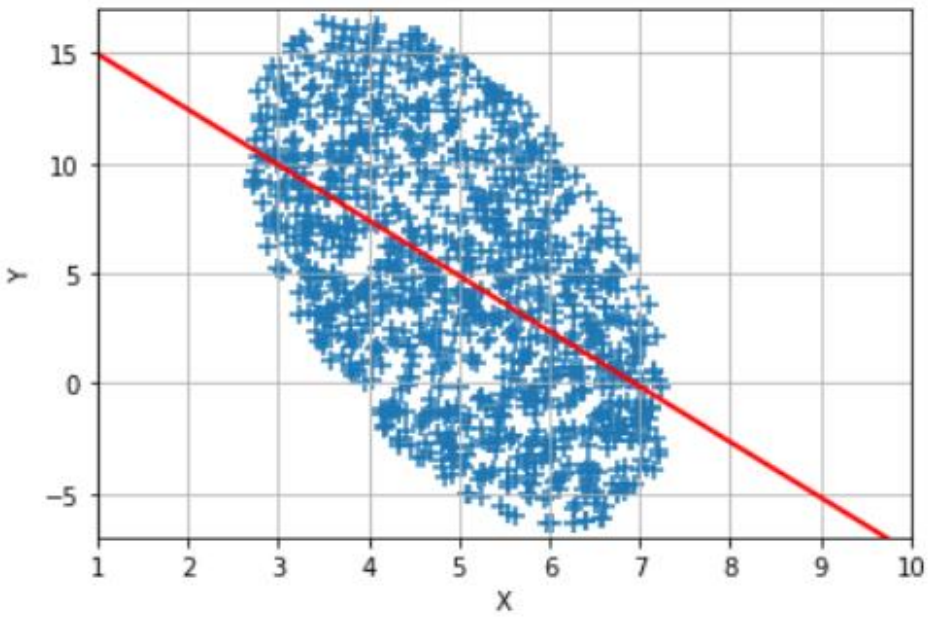
1.7

```
=== Run information ===

Scheme:       weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4
Relation:     XY-(5-5-5-circle)
Instances:    1000
Attributes:   2
              X
              Y
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===


Linear Regression Model

Y =

 +
      5.0287

Time taken to build model: 0 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                 -0.0883
Mean absolute error                      2.0946
Root mean squared error                  2.4657
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances             1000
```



1.8

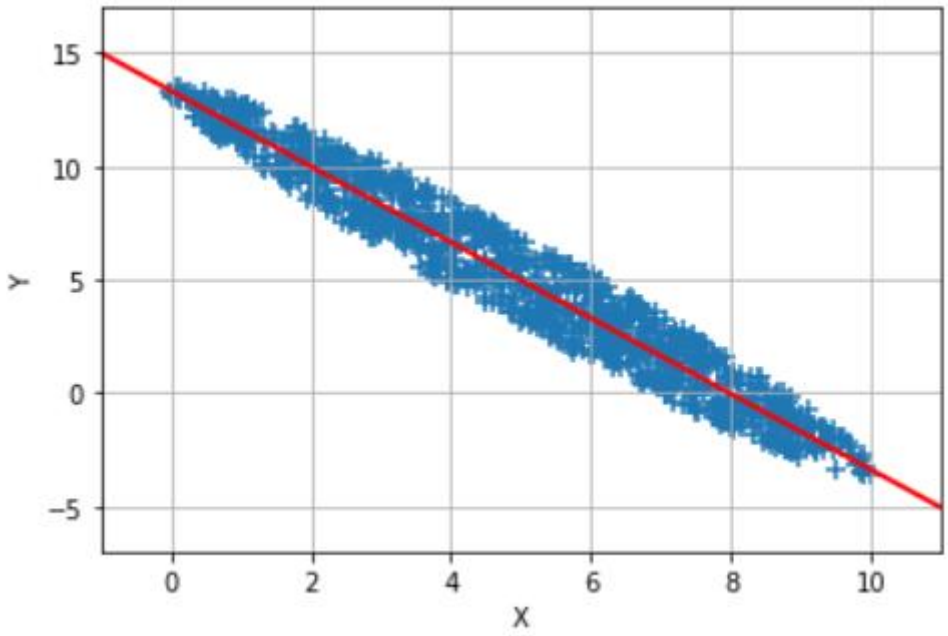
```
=== Run information ===

Scheme:       weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4
Relation:     XY-(fat-ellipse)
Instances:    1000
Attributes:   2
              X
              Y
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===


Linear Regression Model

Y =

     -2.5216 * X +
     17.4559

Time taken to build model: 0 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.5149
Mean absolute error                      4.1194
Root mean squared error                  4.9034
Relative absolute error                 85.3231 %
Root relative squared error             85.6643 %
Total Number of Instances             1000
```



Krzywa regresji nie przechodzi przez środek elipsy, ponieważ prosta jest wyznaczana na podstawie odległości punktów nad oraz pod nią, z których liczona jest średnia. W przypadku elipsy równoległo umieszczonej do podłoża – krzywa regresji przechodziłaby przez środek.

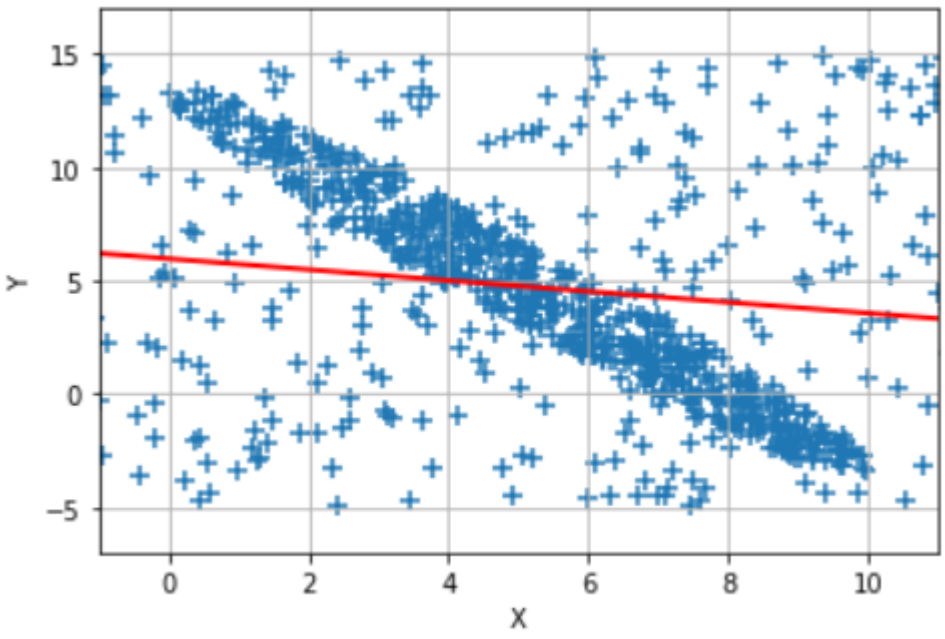
1.8

```
=== Run information ===

Scheme:       weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4
Relation:     XY-(ellipse)
Instances:    1000
Attributes:   2
              X
              Y
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===


Linear Regression Model

Y =

     -1.6706 * X +
     13.3334

Time taken to build model: 0 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.975
Mean absolute error                      0.825
Root mean squared error                  0.9704
Relative absolute error                 22.1431 %
Root relative squared error             22.1937 %
Total Number of Instances             1000
```



1.9

```
=== Run information ===

Scheme:       weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4
Relation:     XY-(ellipse-outliers)
Instances:    1000
Attributes:   2
              X
              Y
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===


Linear Regression Model

Y =

     -0.2412 * X +
      5.974

Time taken to build model: 0 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.186
Mean absolute error                      4.1123
Root mean squared error                  4.9593
Relative absolute error                 96.1878 %
Root relative squared error             98.117  %
Total Number of Instances             1000
```



1.10

```
[1]  import matplotlib.pyplot as plt
     import numpy as np
     from scipy import stats
     from sklearn import linear_model
     ✓ 6.4s

[3]  inp = "dane/xy-002.arff"
     x, y = np.loadtxt(inp, delimiter=',', usecols=(0, 1), unpack=True,skiprows=6)
     ✓ 0.0s

[4]  features=x.reshape(x.size,1)
     ✓ 0.0s

[5]  regr = linear_model.LinearRegression()
     regr.fit(features, y)
     ✓ 0.1s

...  LinearRegression()
```

```
print('Coefficients: ', regr.coef_, ' Intercept: ',regr.intercept_)
```

[6] ✓ 0.0s

```
Coefficients:  [-74.32713092]  Intercept:  652.8037321131726
```

+ Code + Markdown

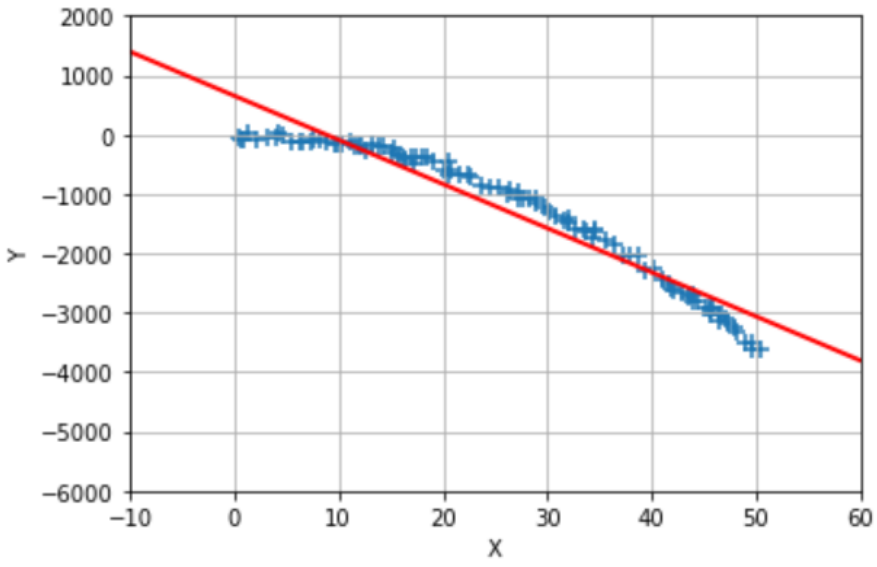
```
fx=np.linspace(-10,60,100)

fy= regr.coef_[0]* fx + regr.intercept_
```

[8] ✓ 0.0s

```
plt.plot(fx,fy,linewidth=2,color='r')
plt.scatter(x,y,s=80, marker='+')
plt.xlim(-10,60)
plt.ylim(-6000, 2000)
plt.grid(True)
plt.xlabel('X')
plt.ylabel('Y')
```

[11] ✓ 0.1s

```
Text(0, 0.5, 'Y')
```



1.11

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.933
Model:	OLS	**Adj. R-squared:**	0.933
Method:	Least Squares	**F-statistic:**	1370.
Date:	Sun, 12 Mar 2023	**Prob (F-statistic):**	2.10e-59
Time:	10:43:54	**Log-Likelihood:**	-711.10
No. Observations:	100	**AIC:**	1426.
Df Residuals:	98	**BIC:**	1431.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	652.8037	60.627	10.768	0.000	532.491	773.116
x1	-74.3271	2.008	-37.012	0.000	-78.312	-70.342

Omnibus:	11.135	**Durbin-Watson:**	0.062
Prob(Omnibus):	0.004	**Jarque-Bera (JB):**	7.706
Skew:	-0.547	**Prob(JB):**	0.0212
Kurtosis:	2.191	**Cond. No.**	61.2

Standardowy błąd oszacowania współczynników wynosi 2.008 dla x1 oraz 60.627 na stałej.

Wartości mieszczą się w zakresie od -78.312 do -70.342 z 95% wiarygodnością.

Zmienne t są statystykami obliczanymi przy założeniu $H_0 : B_2 = 0$ oraz $H_a : B_2 \neq 0$ i obliczane są ze wzoru $t = (b_1 - B_1) / s.e (b_1)$.

Wartości p to prawdopodobieństwo uzyskania statystyki co najmniej tak sprzecznej z H0, jak obliczono przy założeniu, że hipoteza zerowa jest prawdziwa.

1.12

OLS Regression Results

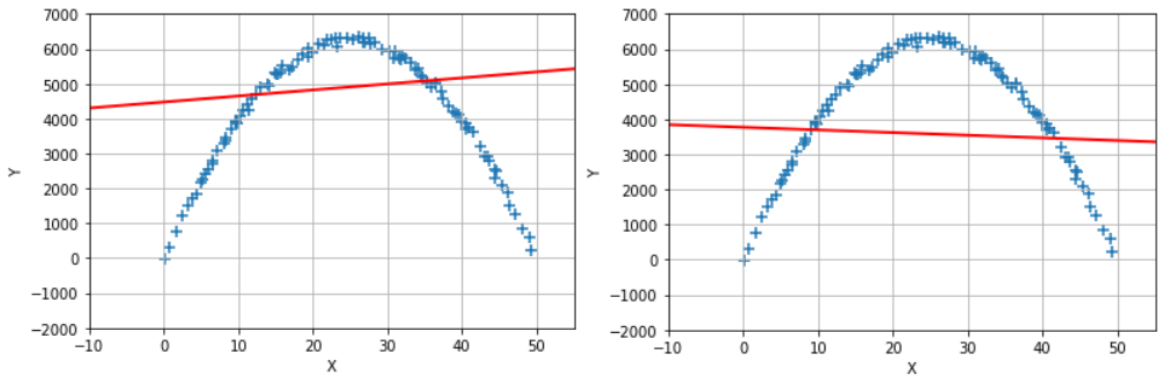
Dep. Variable:	y	**R-squared:**	0.002
Model:	OLS	**Adj. R-squared:**	-0.009
Method:	Least Squares	**F-statistic:**	0.1537
Date:	Sun, 12 Mar 2023	**Prob (F-statistic):**	0.696
Time:	11:11:58	**Log-Likelihood:**	-887.42
No. Observations:	100	**AIC:**	1779.
Df Residuals:	98	**BIC:**	1784.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	4124.0872	352.826	11.689	0.000	3423.916	4824.258
x1	4.8670	12.415	0.392	0.696	-19.771	29.505

Omnibus:	9.293	**Durbin-Watson:**	0.013
Prob(Omnibus):	0.010	**Jarque-Bera (JB):**	8.946
Skew:	-0.671	**Prob(JB):**	0.0114
Kurtosis:	2.414	**Cond. No.**	57.5

Standardowy błąd oszacowania współczynników wynoszą 12.415 dla x1 oraz 352.826 dla stałej.

Wartości mieszczą się w zakresie od -19.771 do 29.505 z 95% wiarygodnością.



Rysunek 1 Skrajne przebiegi