Home Work1

Aidar Ahmetshin

February 2020

1 Task 1

1.1 Outliers

To exclude outliers from our time serious data, we should use 2 sigma rule.
We calculate the 1) mean value of our data 2) standard deviation and then remove points that doesn't exists in 2 sigma range
Our dataset has uniform distribution. Find confidence interval with confidence level = 0.95 for known mean and standard deviation.

$$CI = mean + confidencelevel(0.95) * \frac{standarddeviation}{\sqrt{n_elements}} \tag{1}$$

Plot histogram of values distribution and put bounds of confidence intervals.

1.2 Linear Regression

The equation of regression line is represented as:

$$y(x_i) = b_0 + b_1 * x_i \tag{2}$$

y(x) represents the predicted response value for i th observation. b0 and b1 are regression coefficients and represent y-intercept and slope of regression line respectively.
To create our model, we must "learn" or estimate the values of regression coefficients b0 and b1. And once we should estimated these coefficients via Least Squares technique.

$$y_i = b_0 + b_1 * x_i + \epsilon_i = y(x_i) + \epsilon_i => \epsilon_i = y_i - y(x_i) \tag{3}$$

Where ϵ_i is residual error on each i th step. So our goal to minimize the total residual error. We need to get equation for b_0 and b_1 of our model from:

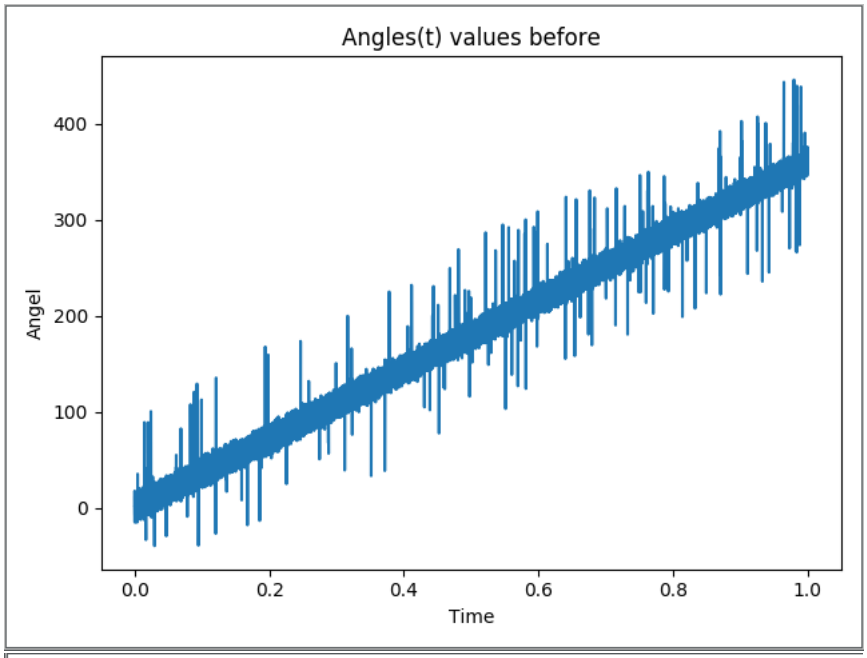
$$J(b_0, b_1) = \frac{1}{2 * n} * \sum \epsilon_i^2 \tag{4}$$

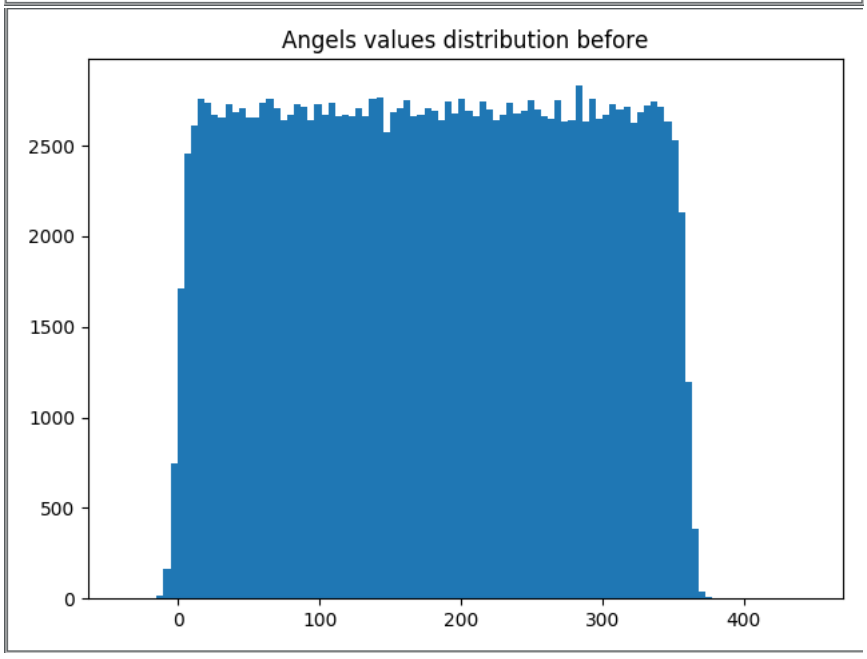
The result equations will be look like this:

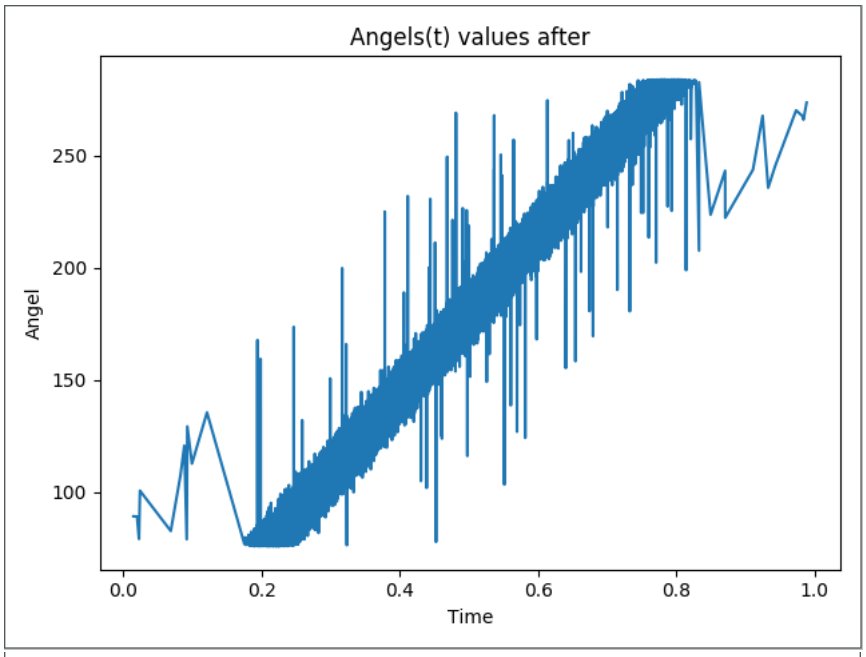
$$b_1 = \frac{sumofcross - deviationsofyandx}{sumofsquareddeviationsofx} \tag{5}$$

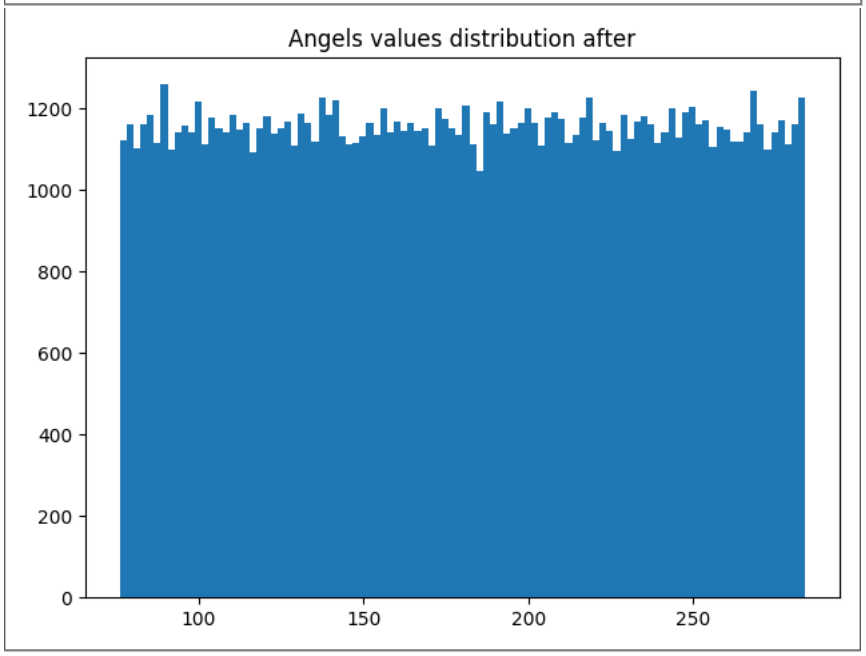
$$b_0 = y(mean) - b_1 * x(mean) \tag{6}$$

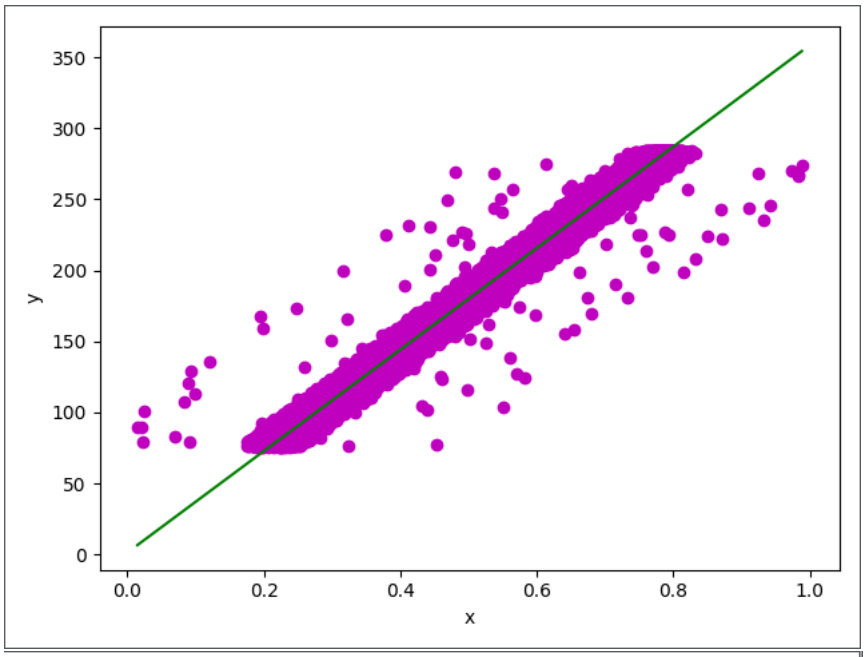
1.3 Results

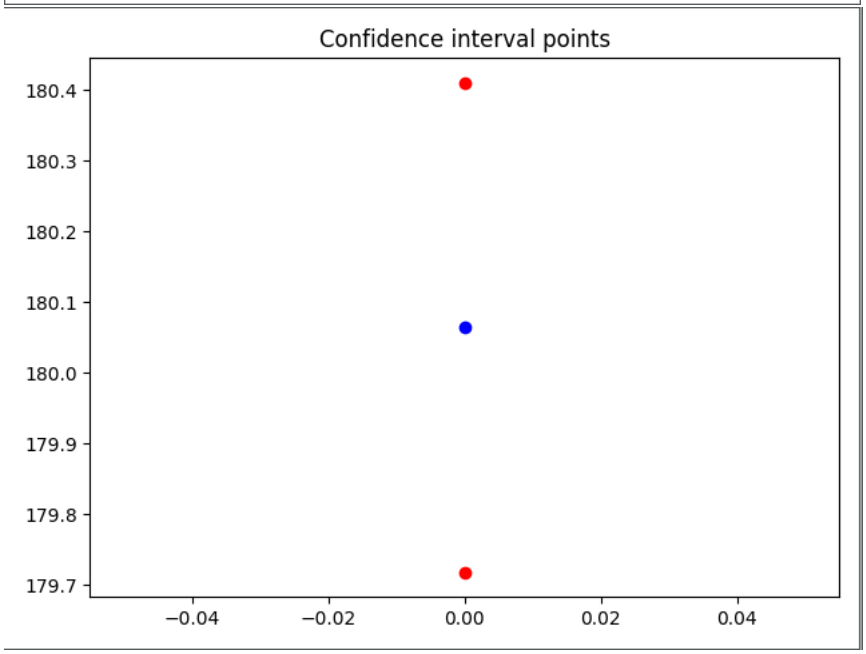

Angles(t) values before
Angel
Time
0
100
200
300
400
0.0
0.2
0.4
0.6
0.8
1.0

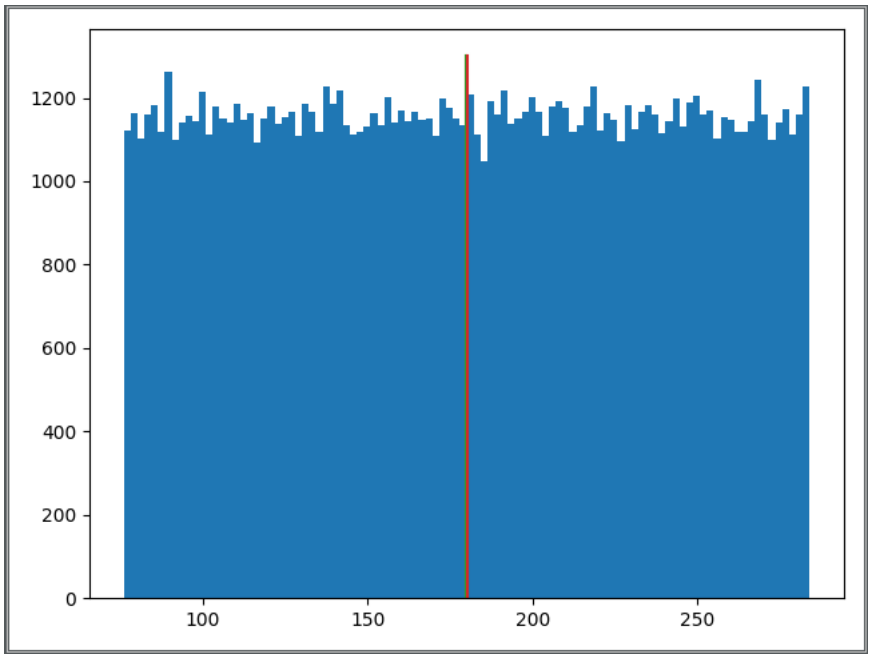

Angels values distribution before
0
500
1000
1500
2000
2500
0
100
200
300
400


Angels(t) values after
Angel
Time
250
200
150
100
0.0
0.2
0.4
0.6
0.8
1.0


Angels values distribution after
1200
1000
800
600
400
200
0
100
150
200
250


350
300
250
200
150
100
50
0
y
0.0
0.2
0.4
0.6
0.8
1.0
x


Confidence interval points
180.4
180.3
180.2
180.1
180.0
179.9
179.8
179.7
−0.04
−0.02
0.00
0.02
0.04



2 Task 2

2.1 RANSAC

The main principle we have n numbers of iterations during which the algorithm take m points for generate plane with some interval choose points that exists there and take it like inliers points. To do it we should defined threshold. Next step is evaluate model of choosen points as new model of the plane, so this iterative algorithm at the end give us the best fitted model for current data.
Select randomly the minimum number of points required to determine the model parameters.
Solve for the parameters of the model.
Determine how many points from the set of all points fit with a predefined plane.
If the fraction of the number of inliers over the total number points in the set exceeds a predefined threshold, re-estimate the model parameters using all the identified inliers and terminate.
Let u represent the probability that any selected data point is an inlier and $v = 1u$ the probability of observing an outlier. N iterations of the minimum number of points denoted m are required, where:

$$1 - p = (1 - u^m)^N \tag{7}$$

$$N = \frac{log(1-p)}{log(1-(1-v)^m)} \tag{8}$$

2.2 Results