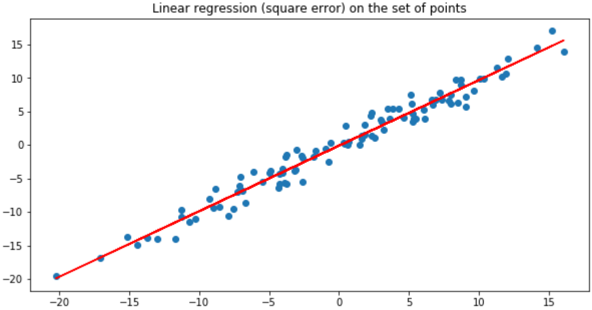
Linear regression using robust functions

Lluís Garrido – lluis.garrido@ub.edu

Master in Data Science – UB

Linear regression using squared error

Assume that we have a set of points we would like to approximate using a line. How can we compute the line?



Linear regression using squared error

One way to tackle this problem is as follows.

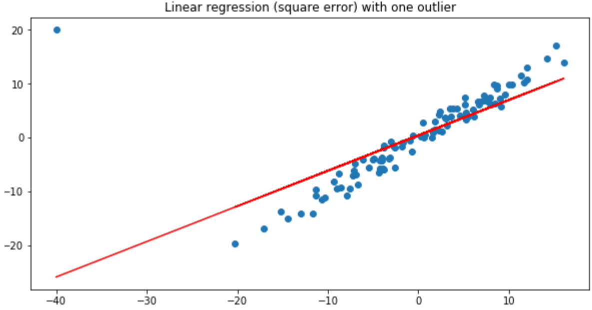
1. Let $\mathbf{x}_i = \{x_i, y_i\}$ with $i = 1 \dots m$ be the data points.
2. Assume we model the best fitting line as $\hat{y}_i = w_0 x_i + w_1$. We have to find the optimum values of w_0 and w_1.
3. To find such "good" values we have to model the error between y_i and $\hat{y}_i$. A simple way to proceed is to use the squared error

$$Q(w_0, w_1) = \frac{1}{2} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2$$

4. Applying gradient descent we may find the optimum values w_0 and w_1.

Linear regression using squared error

The squared error works fine if all the data points follow a Gaussian model. When working with real data, there may be some points that do not follow the expected model. What happens if we have "outliers", i.e. data that does not follow our expected model?



A single "outlier" may have a large influence in the best fitting line!

Linear regression using squared error

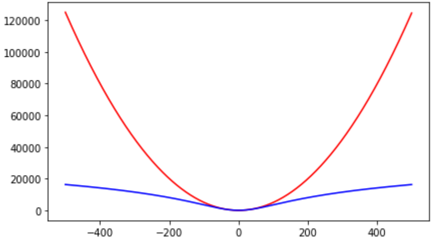
This is the fitting line, using squared error, if two outliers are present!

Linear regression using robust functions

How can we fit a line being “robust” to outliers? One way to proceed is to use robust functions in the error function (rather than a squared error).

$$Q(w_0, w_1) = \sum_{i=1}^{m} \rho(e_i) \qquad e_i = |\hat{y}_i - y_i|$$

where $\rho(.)$ is a robust functions. E.g. red, the squared error; blue, the Cauchy robust function.



Linear regression using robust functions

This is the result for linear regression using the Cauchy robust function (used in the lab)

In the courses of Machine Learning and Deep Learning you'll notice that "robust" function are applied to be robust to outliers! There are many available robust functions, just use the one that fits your needs.