**Least-Squared Linear Regression**

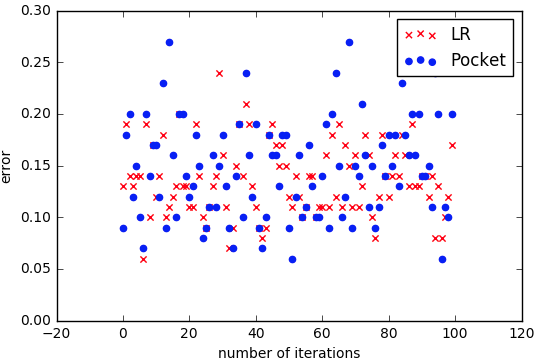
Theory:

We define the line that best fits the data by trying to minimize the distance between each data point and the line we fit.  This is done by minimizing the error function that minimizes the sum of all distances between data points and the fitted line.

The most common error function used is the **least squared optimization** which chooses the parameters in order to minimize the squared difference between the prediction (line) and actual data value summed over all of the data points.

Main Part:

We are given non-separable data points. First, we have to solve classification problem with Pocket Algorithm (almost the same, but we use the best weight for all iterations), where we find a line that separates out the classes so that they can be distinguished. Then we solve Linear Regression problem, where we fit a line to the data. Here we apply pseudo-inverse technique to get the analytical answers.



Mean for E(w\_lin:) 0.1359

Mean for E(w\_pocket:) 0.1462

So we get here that Linear Regression performs slightly better than Pocket Algorithm.