# EM for LS

Given a set of data samples , where , we desire to fit the relation , which is trained by minimizing the least-squared error.

The quality of each data sample are not consistent, and there may exist anomaly in the data samples. We would like to assign different importance to different data samples, which can be implemented by assigning weights in regression loss.

We assume , such that the error of each data sample yields a Gaussian distribution with distinct error. Then the data sample follows the distribution . We desire to optimize the likelihood

Maximize the logarithm of is equivalent to minimizing , thus,

样本来自三个不同的数据集，其标准差就三个，取值就三个，隐变量迭代的过程就是将样本正确分类的过程，但是在不同单个样本分类过程中三个数据集的标准差会发生变化。

EM求解：

初始化，

E：求解隐变量，已知样本数据，模型系数，求解

M：更新模型系数，已知标准差，样本数据，求解

实验：

'''

更新隐变量

公式min(ai(y-Xw))判断使用哪个的方差最小就将该行归属于哪一类（三批）

将与x和y相乘直接利用相乘之后的数值进行ls的求解

'''

# EM for LS

Given a set of data samples , where , we desire to fit the relation , which is trained by minimizing the least-squared error.

The quality of each data sample are not consistent, and there may exist anomaly in the data samples. We would like to assign different importance to different data samples, which can be implemented by assigning weights in regression loss.

How to compute ? We assume , such that the error of each data sample yields a Gaussian distribution with distinct error. Then the data sample follows the distribution . We desire to optimize the likelihood

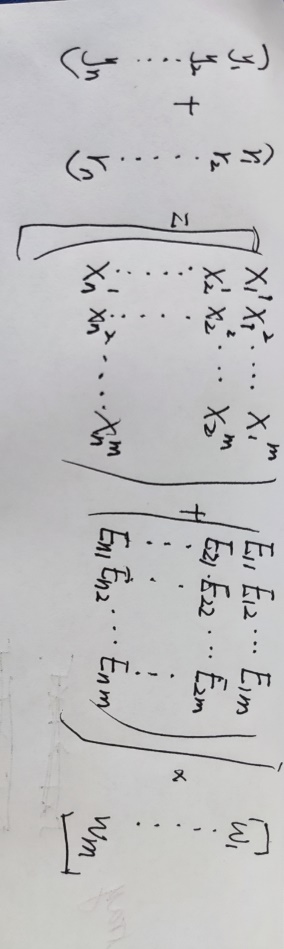
Maximize the logarithm of is equivalent to minimizing , thus,

As the parameters of distribution cannot be known in advance, we need to optimize the likelihood in a EM fashion while considering as a hidden variable.

Given an initial group of model coefficients (can be computed using OLS), we can estimate the variances of each data sample and update the weight for next iteration.

How to compute ? Since it cannot be estimated using a single data sample, we can run a clustering method to group the data samples (based on the residual of each data sample) and those in the same group are assigned the same variance. Or we can start from a unified variance and the variance of each data sample is adapted in each iteration.

TLS按照行：



In the setup of TLS, we desire to fit the relation , and we let , is the estimated value of . We consider both the error for estimating as and the error for measuring as . We assume and ,follow the Gaussian distribution , the probability of data sample conditioned on the model is , where and .

Thus, the likelihood is :

Similarly, we can assume that and follow . Thus, the likelihood is updated to

Maximize the log-likelihood is equivalent to minimize

Given a , use KKT to compute , . We can thus estimate for each data sample and use it to optimize in next iteration.

We assume , follow the Gaussian distribution , and follows the Gaussian distribution ,the probability of data sample conditioned on the model is , where and .

Thus, the likelihood is :

, s.t.

Let,

, s.t.

,

KKT Condition: (1), (2), (3)

From (1) and (2): , , (4)

From (3) and (4): , , (5)

From (4) and (5): (6)

Put (5) and (6) into the objective function:

-> ->

Compute by solving the standard TLS problem, and then compute

Initialize and

E: Estimate and , compute (该如何计算)

M: Compute based on given