# 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.

# EM for TLS

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 , the -th element of , 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. Given , how to compute ? Just scale each data sample by multiplying both and to

# Different Noise Level on Each Column

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 :

Maximize this likelihood is equivalent to minimize

The corresponding optimization problem is

, s.t.

We consider two scenarios:

## When the variances of all columns in are identical such that

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

M: Compute based on given

## When the variances of all columns in are not assumed to be identical

, s.t.

, s.t.

, s.t. ,

,

KKT:

(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:

Let , the problem is equivalent to

Initialize and

E: Estimate variances , , update

M: Compute based on current

根据E和r，求出，计算误差的过程

根据求出

行列结合：

1.分割数据集：列为单位，不同批次加上不同分布的噪声（行）,先在行上进行加权值，与样本数据相乘之后再根据求解列的方法迭代得到新的模型系数