

This document describes an approach to solve a regression problem in case of noise-signal dependent.

Let us assume the following model for data

$$y_i = ax_i + b + x_i^\gamma u + v, \quad (1)$$

where $u \sim \mathcal{N}(0, \sigma_u^2)$, $v \sim \mathcal{N}(0, \sigma_v^2)$. In this way, $e_i = x_i^\gamma u + v$ and by changing $\gamma, \sigma_u^2, \sigma_v^2$, various types of noise can be modeled. For simplicity, let's set $\gamma = 1$ for the given datasets. Therefore, from now on we assume $e_i = x_i u + v$ and the unknown parameters to be estimated are $a, b, \sigma_u^2, \sigma_v^2$.

Also, conditional mean and variance are

$$\mathbb{E}\{e_i|x_i\} = 0, \quad (2)$$

$$\text{Var}\{e_i|x_i\} = \sigma_i^2 = x_i^2 \sigma_u^2 + \sigma_v^2. \quad (3)$$

We can easily see that

$$\mathbb{E}\{y_i|x_i\} = ax_i + b \quad (4)$$

$$\text{Var}\{y_i|x_i\} = \sigma_i^2 \quad (5)$$

$$p(y_i|x_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(y_i - ax_i - b)^2}{2\sigma_i^2}\right). \quad (6)$$

Since samples are assumed independent, we can write

$$p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^N p(y_i|x_i) \quad (7)$$

Taking the logarithm from the above gives

$$\log p(\mathbf{y}|\mathbf{x}) = \sum_{i=1}^N \log p(y_i|x_i) = -\frac{N}{2} \log 2\pi - \frac{1}{2} \sum_{i=1}^N \log \sigma_i^2 - \frac{1}{2} \sum_{i=1}^N \frac{(y_i - ax_i - b)^2}{\sigma_i^2} \quad (8)$$

Now for maximum likelihood estimation of the unknown parameters, we maximize $p(\mathbf{y}|\mathbf{x})$ or equivalently minimize the following cost function

$$J(a, b, \sigma_u^2, \sigma_v^2) = \frac{1}{2} \sum_{i=1}^N \log \sigma_i^2 + \frac{1}{2} \sum_{i=1}^N \frac{(y_i - ax_i - b)^2}{\sigma_i^2} \quad (9)$$

The derivatives of the the cost function respect to parameters are

$$\frac{\partial J}{\partial a} = \sum_{i=1}^N \frac{-x_i(y_i - ax_i - b)}{\sigma_i^2}, \quad (10)$$

$$\frac{\partial J}{\partial b} = \sum_{i=1}^N \frac{-(y_i - ax_i - b)}{\sigma_i^2}, \quad (11)$$

$$\frac{\partial J}{\partial \sigma_u^2} = \frac{1}{2} \sum_{i=1}^N \frac{x_i^2}{\sigma_i^2} - \frac{1}{2} \sum_{i=1}^N \frac{x_i^2(y_i - ax_i - b)^2}{(\sigma_i^2)^2}, \quad (12)$$

$$\frac{\partial J}{\partial \sigma_v^2} = \frac{1}{2} \sum_{i=1}^N \frac{1}{\sigma_i^2} - \frac{1}{2} \sum_{i=1}^N \frac{(y_i - ax_i - b)^2}{(\sigma_i^2)^2} \quad (13)$$

We can perform a gradient checking by computing the numerical gradient and analytic gradient to make sure that the above equations are correct. The gradient checking coded is included.

Now, we can estimate parameters using the gradient descent as

$$a^{(k)} = a^{(k-1)} - \alpha \frac{\partial J}{\partial a}, \quad (14)$$

$$b^{(k)} = b^{(k-1)} - \alpha \frac{\partial J}{\partial b}, \quad (15)$$

$$\sigma_u^{2(k)} = \sigma_u^{2(k-1)} - \alpha \frac{\partial J}{\partial \sigma_u^2}, \quad (16)$$

$$\sigma_v^{2(k)} = \sigma_v^{2(k-1)} - \alpha \frac{\partial J}{\partial \sigma_v^2}, \quad (17)$$

where α is the learning rate.

To assess the quality of estimated parameters, we can compute the mean and variance of estimated parameters from the maximum likelihood criterion. If we set Eq. (10) to zero we have

$$\hat{a} = \frac{\sum_{i=1}^N \frac{x_i(y_i - b)}{\sigma_i^2}}{\sum_{i=1}^N \frac{x_i^2}{\sigma_i^2}} \quad (18)$$

It can be seen that

$$\mathbb{E}\{\hat{a}|x_i, \sigma_u^2, \sigma_v^2\} = a \quad (19)$$

$$\text{Var}\{\hat{a}|x_i, \sigma_u^2, \sigma_v^2\} = \frac{1}{\sum_{i=1}^N \frac{x_i^2}{\sigma_i^2}} \quad (20)$$

Similarly for b we have

$$\hat{b} = \frac{\sum_{i=1}^N \frac{(y_i - ax_i)}{\sigma_i^2}}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} \quad (21)$$

$$\mathbb{E}\{\hat{b}|x_i, \sigma_u^2, \sigma_v^2\} = b \quad (22)$$

$$\text{Var}\{\hat{b}|x_i, \sigma_u^2, \sigma_v^2\} = \frac{1}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} \quad (23)$$

This shows that the estimators are unbiased with some variance that depends on the number of samples, noise variance and also signal values. With increasing the number of samples, the variance of estimates goes to zero.

The scatter plot and linear model for data1.csv file is shown in Fig. 1. The estimated parameters for this data are $a = 0.98, b = 0.05, \sigma_u^2 = 3.6, \sigma_v^2 = 1.1$. Also, based on the estimated values, variance of \hat{a} and \hat{b} are 0.044 and 0.064, respectively.

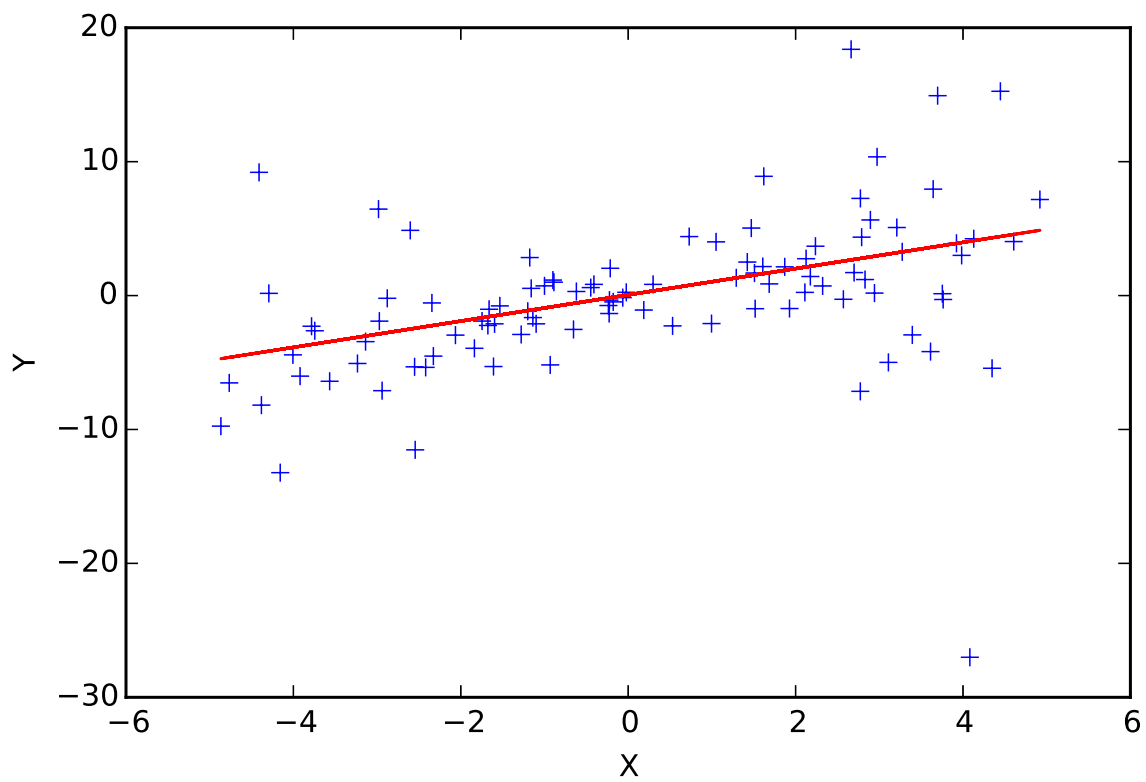


Figure 1: Scatter of data and the fitted linear model.