How to Best Process Data If We Have Both Absolute and Relative Measurement Errors: A Pedagogical Comment

Ana Maria Hernandez Posada[1], Maria Isabel Olivarez[1], Christian Servin[2], and Vladik Kreinovich[1]
[1]Department of Computer Science
University of Texas at El Paso, El Paso, TX 79968, USA
Computer Science and IT Program, El Paso Community College
amhernandezegias@miners.utep.edu, miolivares@miners.utep.edu
servin1@epcc.edu, vladik@utep.edu

1. Formulation of the Problem

- In many practical situations, we need to find the dependence of a quantity y on quantities $x = (x_1, \ldots, x_n)$.
- Usually, we know the type of the dependence, i.e., we know that $f = f(p, x)$ for some parameters

$$p = (p_1, \ldots, p_m).$$

- We just need to find p.
- For example, the dependence may be linear, then

$$f(x, p) = \sum_{i=1}^{n} p_i \cdot x_i + p_{n+1}.$$

- To find this dependence, we measure x_i and y in several situations k.
- Then, we find p for which $f(p, x^{(k)}) \approx y^{(k)}$ for all k.

2. Formulation of the Problem (cont-d)

- The measurement error is often caused by a large number of independent factors of about the same size,
- In this case the Central Limit Theorem implies that it is normally distributed.
- Usually, it is assumed that the bias is 0, so we only have standard deviation σ.
- Sometimes, we have absolute error $\sigma = \text{const}$, in which case we use the usual Least Squares method

$$\sum_k (y^{(k)} - f(p, x^{(k)}))^2 \to \min .$$

- In other cases, we have relative error, in which case we find p for which $\sum_k \frac{(y^{(k)} - f(p, x^{(k)}))^2}{(y^{(k)})^2} \to \min$.

3. Formulation of the Problem (cont-d)

- In practice, we usually have both absolute and relative error components.
- Namely, $\Delta y = \Delta y_{\rm abs} + \Delta y_{\rm rel}$, with $\sigma_{\rm abs} = \sigma_0$ and $\sigma_{\rm rel} = \sigma_1 \cdot |y|$ for some σ_i.
- How should we then process data?

4. Recommendation

- In this case, the variance of the measurement error if $\sigma^2 = \sigma_0^2 + \sigma_1^2 \cdot y^2$.
- So, we use Maximum Likelihood method and maximize the expression

$$\prod_k \frac{1}{\sqrt{2\pi} \cdot \sqrt{\sigma_0^2 + \sigma_1^2 \cdot (y^{(k)})^2}} \cdot \exp\left(-\frac{(y^{(k)} - f(p, x^{(k)}))^2}{2(\sigma_0^2 + \sigma_1^2 \cdot (y^{(k)})^2)}\right).$$

- In this talk, we present an iterative algorithm for finding p.

5. Algorithm

- The above problem is complex, so what we can do is solve it iteratively.
- First, we assume that $\sigma_1 = 0$.
- Then, we compute $(\sigma^{(k)})^2 = \sigma_0^2 + \sigma_1^2 \cdot (y^{(k)})^2$.
- After that, we use the Least Squares and find p that minimizes $\sum\limits_k \dfrac{(y^{(k)} - f(p, x^{(k)}))^2}{(y^{(k)})^2}$.
- Once we find these values p, we again use the Least Squares to find the values σ_0^2 and σ_1^2 for which

$$(y^{(k)} - f(p, x^{(k)}))^2 \approx \sigma_0^2 + \sigma_1^2 \cdot (y^{(k)})^2.$$

- Then, we again compute $(\sigma^{(k)})^2$, find p, etc., until the process converges.