Spring 2020 - Knowledge Discovery in Data at Scale Technologies - Homework #4

by Almaz Murzabekov

Homework description

Optimal Estimation

Consider the following series of measurements of the unknown value x:

$$y_i = a_i x + \varepsilon_i, \quad i = 1, \ldots, n,$$

where y_i are measurement results, a_i are known coefficients, and ε_i represent random error of measurement and are independent identically distributed (i.i.d.) with zero mean and variance σ^2:

$$E\varepsilon_i = 0, E\varepsilon_i^2 = \sigma^2, \quad i = 1, \ldots, n$$

a) What function of $y_1, \ldots, y_n$ (and of $a_1, \ldots, a_n$) would you use as a good estimate $\widehat{x}$ for x? $\widehat{x}$ = ?

(Hint: try best linear estimate or least squares estimate).

Estimation of x might be formulated in the following formula, and the value of this formula we should minimize

$$Q(x) = \sum_{i=1}^{n} \frac{(y_i - a_i x)^2}{\sigma^2} \rightarrow min$$

$$\frac{dQ}{dx} = \sum_{i=1}^{n} \frac{(-2y_i a_i + 2a_i^2 x)}{\sigma^2} = \sum_{i=1}^{n} \frac{-2(y_i a_i - a_i^2 x)}{\sigma^2} = -2\sum_{i=1}^{n} \frac{y_i a_i - a_i^2 x}{\sigma^2} = 0 \Rightarrow$$

$$-2\sum_{i=1}^{n} \frac{y_i a_i - a_i^2 x}{\sigma^2} = 0 \Rightarrow$$

$$-2(\sum_{i=1}^{n} \frac{y_i a_i}{\sigma^2} - \sum_{i=1}^{n} \frac{a_i^2 x}{\sigma^2}) = 0 \Rightarrow$$

$$-2 \neq 0 \Rightarrow (\sum_{i=1}^{n} \frac{y_i a_i}{\sigma^2} - \sum_{i=1}^{n} \frac{a_i^2 x}{\sigma^2}) = 0 \Rightarrow$$

$$(\sum_{i=1}^{n} \frac{y_i a_i}{\sigma^2} - \sum_{i=1}^{n} \frac{a_i^2 x}{\sigma^2}) = 0 \Rightarrow$$

$$\sum_{i=1}^{n} \frac{y_i a_i}{\sigma^2} = \sum_{i=1}^{n} \frac{a_i^2 x}{\sigma^2} \Rightarrow (*\sigma^2)$$

$$\sum_{i=1}^{n} y_i a_i = \sum_{i=1}^{n} a_i^2 x \Rightarrow$$

$$\sum_{i=1}^{n} y_i a_i = x \sum_{i=1}^{n} a_i^2 \Rightarrow$$

$$x \sum_{i=1}^{n} a_i^2 = \sum_{i=1}^{n} y_i a_i \Rightarrow$$

$$\hat{x} = \frac{\sum_{i=1}^{n} y_i a_i}{\sum_{i=1}^{n} a_i^2}$$

b) Is this estimate optimal in any sense?

In this work I'm using the method **Least squares estimate.** This method/function approximation applies the principle of least squares to function approximation, by means of a weighted sum of other functions. The best approximation can be defined as that which minimises the difference between the original function and the approximation; for a least-squares approach the quality of the approximation is measured in terms of the squared differences between the two.

c) Is it a biased or an unbiased estimate?

$$\hat{x} = \frac{\sum_{i=1}^{n} y_i a_i}{\sum_{i=1}^{n} a_i^2} \rightarrow E(\hat{x}) =?$$

$$E(\hat{x}) = E(\frac{\sum_{i=1}^{n} y_i a_i}{\sum_{i=1}^{n} a_i^2}) \Rightarrow \quad where \quad y_i = a_i x + \varepsilon_i,$$

$$E(\hat{x}) = E(\frac{\sum_{i=1}^{n} a_i (a_i x + \varepsilon_i)}{\sum_{i=1}^{n} a_i^2}) \Rightarrow$$

$$E(\hat{x}) = E(\frac{\sum_{i=1}^{n} a_i^2 x + a_i \varepsilon_i}{\sum_{i=1}^{n} a_i^2}) \Rightarrow \quad E(\varepsilon_i) = 0$$

$$E(\hat{x}) = E(\frac{\sum_{i=1}^{n} a_i^2 x}{\sum_{i=1}^{n} a_i^2}) \Rightarrow \quad \text{where } x \text{ - is constant}$$

$$E(\hat{x}) = E(\frac{x \sum_{i=1}^{n} a_i^2}{\sum_{i=1}^{n} a_i^2}) \Rightarrow$$

$$E(\hat{x}) = E(x) = x$$

that means, this is **unbiased** estimation

d) What is its variance (expressed through σ^2)? $Var(\hat{x}) = ?$

$$Var(\hat{x}) = E(\hat{x} - x)^2 \Rightarrow$$

$$Var(\hat{x}) = E(\frac{\sum_{i=1}^{n} y_i a_i}{\sum_{i=1}^{n} a_i^2} - x)^2 \Rightarrow \quad where \quad y_i = a_i x + \varepsilon_i,$$

$$Var(\hat{x}) = E(\frac{\sum_{i=1}^{n} a_i (a_i x + \varepsilon_i)}{\sum_{i=1}^{n} a_i^2} - x)^2 \Rightarrow$$

$$Var(\hat{x}) = E(\frac{\sum_{i=1}^{n} a_i^2 x + a_i \varepsilon_i)}{\sum_{i=1}^{n} a_i^2} - x)^2 \Rightarrow$$

$$Var(\widehat{x}) = E(\frac{\sum_{i=1}^n a_i^2 x + \sum_{i=1}^n a_i \varepsilon_i}{\sum_{i=1}^n a_i^2} - x)^2 \Rightarrow$$

$$Var(\widehat{x}) = E(\frac{\sum_{i=1}^n a_i^2 x}{\sum_{i=1}^n a_i^2} + \frac{\sum_{i=1}^n a_i \varepsilon_i}{\sum_{i=1}^n a_i^2} - x)^2 \Rightarrow$$

$$Var(\widehat{x}) = E(x + \frac{\sum_{i=1}^n a_i \varepsilon_i}{\sum_{i=1}^n a_i^2} - x)^2 \Rightarrow$$

$$Var(\widehat{x}) = E(\frac{\sum_{i=1}^n a_i \varepsilon_i}{\sum_{i=1}^n a_i^2})^2 \Rightarrow$$

$$Var(\widehat{x}) = E(\frac{\sum_{i=1}^n (a_i \varepsilon_i)^2}{(\sum_{i=1}^n a_i^2)^2})) \Rightarrow$$

$$Var(\widehat{x}) = E(\frac{\sum_{i,j=1}^n a_i a_j \varepsilon_i \varepsilon_j}{(\sum_{i=1}^n a_i^2)^2}) \Rightarrow$$

$$Var(\widehat{x}) = \frac{\sum_{i,j=1}^n a_i a_j E(\varepsilon_i \varepsilon_j)}{(\sum_{i=1}^n a_i^2)^2}) \Rightarrow$$

$$E\widehat{x} = \begin{cases} 0 & \text{if } i \neq j \\ \sigma^2 & \text{if } i = j \end{cases}$$

$$Var(\widehat{x}) = \frac{\sigma^2 \sum_i^n a_i^2}{(\sum_{i=1}^n a_i^2)^2} \Rightarrow$$

$$Var(\widehat{x}) = \frac{\sigma^2}{\sum_{i=1}^n a_i^2}$$

e) How would you estimate σ^2 if it is unknown? $\widehat{\sigma^2}$ = ?

$$Var(\varepsilon) = E[(\varepsilon - E(\varepsilon))^2] \quad E(\varepsilon) = 0, \Rightarrow$$

$$Var(\varepsilon) = E[(\varepsilon)^2] = \sigma^2 \Rightarrow$$

$$E(\sum_{i=1}^{n}(y_i - a_i x)^2 = \quad \text{where } y_i = a_i x + \varepsilon_i$$

$$E\sum_{i=1}^{n}(a_i x + \varepsilon_i - a_i x)^2 = E\sum_{i=1}^{n}(\varepsilon_i)^2 = n\sigma^2$$

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n}(y_i - a_i\hat{x})^2}{n} \Rightarrow$$

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n}(y_i^2 - 2a_i y_i \hat{x} + a_i^2 \hat{x}^2)}{n} \Rightarrow$$

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n} y_i^2 - \sum_{i=1}^{n} 2a_i y_i \hat{x} + \sum_{i=1}^{n} a_i^2 \hat{x}^2)}{n} \Rightarrow$$

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n} y_i^2 - 2\hat{x}\sum_{i=1}^{n} a_i y_i + \hat{x}^2 \sum_{i=1}^{n} a_i^2)}{n} \Rightarrow \quad \text{considering that } \hat{x} = \frac{\sum_{i=1}^{n} y_i a_i}{\sum_{i=1}^{n} a_i^2}$$

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n} y_i^2 - 2\frac{\sum_{i=1}^{n} y_i a_i}{\sum_{i=1}^{n} a_i^2}\sum_{i=1}^{n} a_i y_i + (\frac{\sum_{i=1}^{n} y_i a_i}{\sum_{i=1}^{n} a_i^2})^2 \sum_{i=1}^{n} a_i^2}{n} \Rightarrow$$

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n} y_i^2 - 2\frac{(\sum_{i=1}^{n} y_i a_i)^2}{\sum_{i=1}^{n} a_i^2} + \frac{\sum_{i=1}^{n} (y_i a_i)^2}{\sum_{i=1}^{n} a_i^2}}{n}$$

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n} y_i^2 - \frac{\sum_{i=1}^{n} (y_i a_i)^2}{\sum_{i=1}^{n} a_i^2}}{n} \Rightarrow \text{biased estimation}$$

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n} y_i^2 - \frac{\sum_{i=1}^{n} (y_i a_i)^2}{\sum_{i=1}^{n} a_i^2}}{n-1} \qquad \Rightarrow \text{unbiased estimation}$$

f) What would you use as an estimate for $Var(\widehat{x})$ if σ^2 is unknown? $\widehat{Var(\widehat{x})} = ?$

$$Var(\widehat{x}) = \frac{\sigma^2}{\sum_{i=1}^{n} a_i^2} \Rightarrow \text{considering that } \sigma^2 =$$

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n} y_i^2 - \frac{\sum_{i=1}^{n} (y_i a_i)^2}{\sum_{i=1}^{n} a_i^2}}{n-1}$$

$$\widehat{Var(\widehat{x})} = \frac{\sum_{i=1}^{n} y_i^2 - \frac{\sum_{i=1}^{n} (y_i a_i)^2}{\sum_{i=1}^{n} a_i^2}}{(n-1)(\sum_{i=1}^{n} a_i^2)}$$

g) Suppose that the variance σ^2 is known. What "canonical information" would be sufficient to extract from the series of records

$$(y_1, a_1), \ldots, (y_n, a_n), \quad i = 1, \ldots, n$$

\ in order to compute the estimate $\widehat{x}$, and its variance $Var(\widehat{x})$?

$$\widehat{x} = \frac{\sum_{i=1}^{n} y_i a_i}{\sum_{i=1}^{n} a_i^2} \qquad Var(\widehat{x}) = \frac{\sigma^2}{\sum_{i=1}^{n} a_i^2}$$

$$T = \sum_{i=1}^{n} a_i y_i; \quad V = \sum_{i=1}^{n} a_i^2;$$

$$\widehat{x} = \frac{T}{V}; \quad Var(\widehat{x}) = \frac{\sigma^2}{V}$$

h) Suppose that the variance σ^2 is NOT known. What "canonical information" would be sufficient to extract from the series of observations in order to compute $\widehat{x}$, $\widehat{\sigma^2}$, and $\widehat{Var(\widehat{x})}$?

$$T = \sum_{i=1}^{n} a_i y_i; \quad V = \sum_{i=1}^{n} a_i^2; \quad Z = \sum_{i=1}^{n} y_i^2;$$

$$Q = \sum_{i=1}^{n} (a_i y_i)^2; \quad N = \text{Number of observation}$$

$$\hat{x} = \frac{\sum_{i=1}^{n} y_i a_i}{\sum_{i=1}^{n} a_i^2} = \frac{T}{V}$$

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n} y_i^2 - \frac{\sum_{i=1}^{n} (y_i a_i)^2}{\sum_{i=1}^{n} a_i^2}}{n-1} = \frac{Z - \frac{Q}{V}}{N-1}$$

$$\widehat{Var(\hat{x})} = \frac{\sum_{i=1}^{n} y_i^2 - \frac{\sum_{i=1}^{n} (y_i a_i)^2}{\sum_{i=1}^{n} a_i^2}}{(n-1)(\sum_{i=1}^{n} a_i^2)} = \frac{Z - \frac{Q}{V}}{(N-1)V}$$

i) How should we update such "information" when a new record (y_{n+1}, a_{n+1}) arrives?

Case when σ^2 is known

$$(T_1, V_1) \oplus (y_{n+1}, a_{n+1}) = (\\ T_1 + y_{n+1} a_{n+1}, \\ V_1 + a_{n+1}^2)$$

Case when σ^2 is unknown

$$(T_1, V_1, Z_1, Q_1, N_1) \oplus (y_{n+1}, a_{n+1}) = (\\ T_1 + y_{n+1} a_{n+1}, \\ V_1 + a_{n+1}^2, \\ Z_1 + y_{n+1}^2, \\ Q_1 + (y_{n+1} a_{n+1})^2, \\ N_1 + 1)$$

j) How should we "combine" (merge) two pieces of "canonical information"?

Case when σ^2 is known

$$(T_1, V_1) \oplus (T_2, V_2) = (T_1 + T_2, V_1 + V_2)$$

Case when σ^2 is unknown

$$(T_1, V_1, Z_1, Q_1, N_1) \oplus (T_2, V_2, Z_2, Q_2, N_2) =$$
$$(T_1 + T_2, V_1 + V_2, Z_1 + Z_2, Q_1 + Q_2, N_1 + N_2)$$