Information in "Raw" form.

$(x_1, x_2 \cdots, x_n)$ mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

* Existence ok
* Uniqueness - no, can change order.
  * Elem $x \mapsto (x)$
  * Empty $\emptyset \mapsto ()$
* Composition

$$(x_1 \cdots x_n) \oplus (y_1, \cdots y_m) =$$
$$= (x_1 \cdots x_n, y_1 \cdots y_m)$$

- Comm NO

$$(y_1 \cdots y_m) \oplus (x_1 \cdots x_n) = (y_1 \cdots x_n).$$

- Assoc $(x_1 \cdots x_n) \oplus (y_1 \cdots y_m) \oplus (z_1 \cdots z_4)$

$$= (x_1 \cdots x_n, y_1 \cdots y_m, z_1 \cdots z_4) \quad \checkmark.$$

- Neutral el-t
* Completeness true
* minimal. NO!
* Efficiency.
  * updating: - simple but lots of data
  * deployment: NO. Each time need to recompute from scratch.

Random variable ν.

$E\nu$ and $\mathrm{Var}\,\nu = E(\nu - E\nu)^2$

$\nu \sim (\mu, \sigma^2)$ $\quad E\nu = \mu, \ \mathrm{Var}\,\nu = \sigma^2$.

$\nu_1 \dots \nu_n \sim (\mu, \sigma^2)$ - independent

Estimate for μ

$\hat{\mu}$ = sample mean of νs

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \nu_i$$

Unbiased! $\quad E\hat{\mu} = \mu$

Est of σ^2? $\quad \underbrace{\frac{1}{n} \sum_{i=1}^{n} (\nu - \hat{\mu})^2}$?

$$E \sum_{i=1}^{n} (\nu_i - \hat{\mu})^2 = E \sum_{i=1}^{n} [\underbrace{\nu_i - \mu} - (\hat{\mu} - \mu)]^2$$

$\varepsilon_i = \nu_i - \mu \quad E\varepsilon_i = E\nu_i - \mu = \mu - \mu = 0$

$$\varepsilon_i \sim (0, \sigma^2) \quad \mathrm{Var}\,\varepsilon_i = E\underbrace{(\varepsilon_i - \underbrace{E\varepsilon_i}_{=0})}_{= \nu_i - \mu}{}^2 = E\varepsilon_i^2 = \sigma^2$$

$$= E \sum_i \Big[\varepsilon_i - \underbrace{\Big(\frac{1}{n} \sum_{j=1}^{n} \vartheta_j - \mu \Big)}_{\frac{1}{n} \sum_j (\vartheta_j - \mu) = \frac{1}{n} \sum_j \varepsilon_j} \Big]^2 =$$

$$= E \sum_{i=1}^{n} \Big[\varepsilon_i - \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j \Big]^2 =$$

$$= \sum_i E \Big[\varepsilon_i^2 - \frac{2}{n} \varepsilon_i \sum_{j=1} \varepsilon_j + \frac{1}{n^2} \sum_j \varepsilon_j \cdot \sum_k \varepsilon_k \Big]$$

$E \varepsilon_i \varepsilon_j = \sigma^2$ if $i = j$

if $i \neq j$ $\quad E \varepsilon_i \varepsilon_j \overset{\text{indep}}{=} \underbrace{E \varepsilon_i}_{=0} \cdot \underbrace{E \varepsilon_j}_{} = 0$

$$= \sum_i \Big[\sigma^2 - \frac{2}{n} \sigma^2 + \frac{1}{n^2} \cdot n \sigma^2 \Big] =$$

$$= n \sigma^2 - 2 \sigma^2 + \underbrace{n \frac{1}{n^2} n}_{=1} \sigma^2 =$$

$$= (n - 2 + 1) \sigma^2 = (n-1) \sigma^2$$

$$\boxed{\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\vartheta_i - \hat{\mu})^2}$$

Suppose that μ - known.

$$\nu_i \sim (\mu, \sigma^2)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum \underbrace{(\nu_i - \mu)}_{\varepsilon_i}^2$$

$$E\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2 = \frac{1}{n} n\sigma^2 = \sigma^2$$

$\Rightarrow$ Unbiased.

Learning

$$x \longrightarrow \boxed{f} \longrightarrow y$$

Applies f and adds noise.

Learning sequence:

$$\left.\begin{array}{l} (x_1, y_1) \\ (x_2, y_2) \\ \vdots \\ (x_n, y_n) \end{array}\right] \longrightarrow \text{info about } f \longrightarrow \hat{f}$$

Known (x_n)
Observed (y_n)

Given: $x \longrightarrow \underset{\text{apply}}{(*)} \longrightarrow \hat{f}(x)$ - est of $f(x)$ and its precision.

Simple Linear Regression.

$(x_1, y_1) \cdots (x_n, y_n)$

$f(x) = \underline{a} + \underline{b}\,x \quad a, b$ - unknowns

$y_i = a + bx_i + \varepsilon_i \quad \varepsilon_i \sim (0, \sigma^2)$

ε_i - independent

Goals!

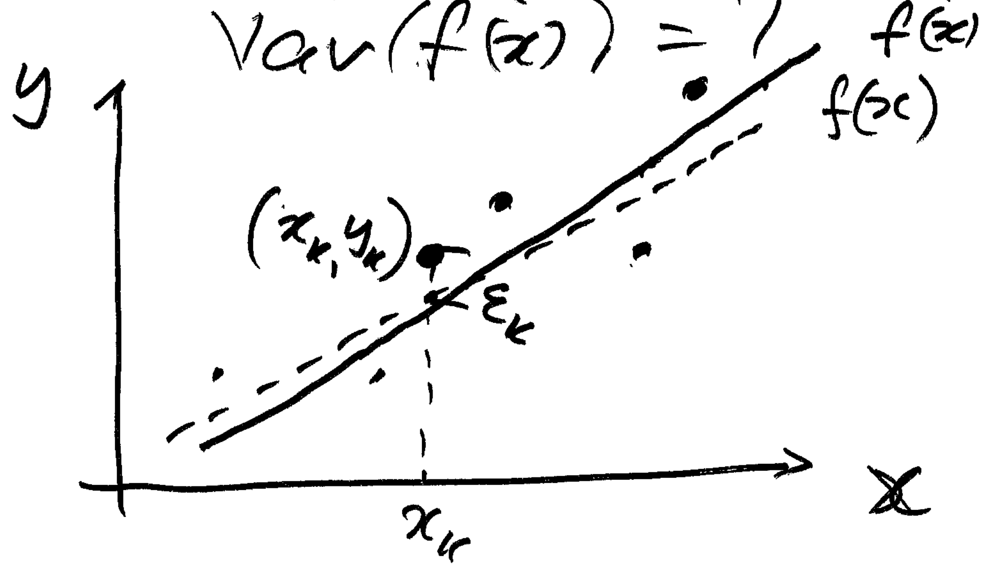
(a) Estimate a and b: $\hat{a}, \hat{b} = ?$

(b) Predict (estimate) $f(x)$ for a given x: $\widehat{f(x)}$ and its precision: $Var(\widehat{f(x)})$

(c) If σ^2 is unknown, est it: $\hat{\sigma}^2 = ?$

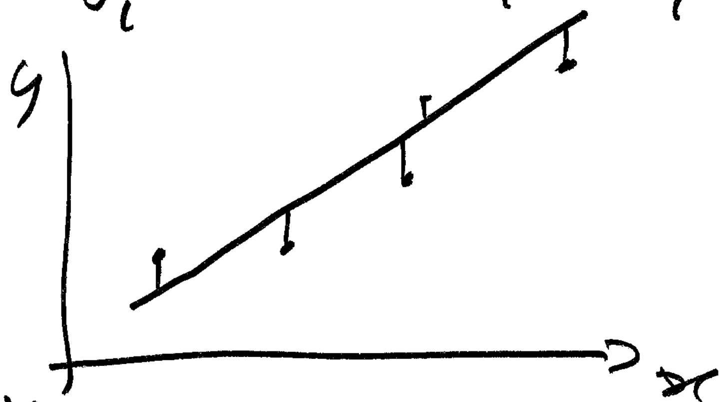
(d) Estimate for $Var(\widehat{f(x)})$

$\widehat{Var}(\widehat{f(x)}) = ?$



Raw Data (x_i, y_i) $i = 1, \dots n$.

$$y_i = a + b x_i + \varepsilon_i$$



$$\sum_{i=1}^{n} \left[y_i - (a + b x_i) \right]^2 = Q(a,b) \sim \min_{a,b}$$

$$Q(a,b) = \sum_i y_i^2 + \sum a^2 + \sum b^2 x_i^2$$

$$- 2 \sum y_i a - 2 \sum y_i b x_i + 2 \sum a b x_i$$

$$= \underbrace{\sum_i y_i^2}_{V} + n a^2 + b^2 \underbrace{\sum x_i^2}_{= U}$$

$$- 2a \underbrace{\sum y_i}_{= Y} - 2 b \underbrace{\sum x_i y_i}_{Z} + 2ab \underbrace{\sum x_i}_{= X}$$

$$= n a^2 + 2 X a b + U b^2 - 2 Y a - 2 Z b + V$$

$$\left.\begin{aligned}\frac{\partial Q}{\partial a} &= 2na + 2Xb - 2Y = 0\\ \frac{\partial Q}{\partial b} &= 2Xa + 2Ub - 2Z = 0\end{aligned}\right\} \Rightarrow \hat{a}, \hat{b}$$

$$na + Xb = Y \qquad \times X$$
$$Xa + Ub = Z \qquad \times n$$

$$(X^2 - nU)\,b = XY - nZ$$

$$\Rightarrow \hat{b} = \frac{Z - \frac{XY}{n}}{U - \frac{X^2}{n}} \qquad \hat{a} = \frac{Y - X\hat{b}}{n}$$

Sample variance for $\{x_i\}$:

$$\overline{Var}(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \qquad \bar{x} = \frac{1}{n}\sum_i x_i \text{ sample mean}$$

$$\ldots = \frac{1}{n-1}\left[\sum_i x_i^2 - \frac{(\sum x_i)^2}{n}\right] =$$

$$= \frac{1}{n-1}\left[U - \frac{X^2}{n}\right]$$

Sample covariance for $\{(x_i, y_i)\}$

$$\overline{Cov}(x, y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$\ldots = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i y_i - \frac{\sum_i x_i \cdot \sum_i y_i}{n}\right] =$$

$$= \frac{1}{n-1}\left[Z - \frac{XY}{n}\right]$$

$$\hat{b} = \frac{\overline{\mathrm{cov}}(x,y)}{\overline{\mathrm{var}}(x)} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$$

$$\hat{a} = \bar{y} - \hat{b}\,\bar{x}$$

$$\left.\begin{matrix}(x_1, y_1)\\ \vdots \\ (x_n\ y_n)\end{matrix}\right] \to (n, X, Y, Z, U)$$

$$X = \sum x_i \quad Y = \sum Y_i$$

$$Z = \sum x_i y_i \quad U = \sum x_i^2$$

$$(x, y) \to \oplus \to (n+1, X+x, Y+y, Z+xy, U+x^2)$$

$$\left.\begin{matrix}(x_1, y_1)\\ \vdots \\ (x_n\ y_n)\end{matrix}\right] \to (n, X, Y, Z, U)$$

$$\oplus \to (n+n', X+X', Y+Y', Z+Z', U+U')$$

$$\left.\begin{matrix}(x_1', y_1')\\ \vdots \\ (x_{n'}', y_{n'}')\end{matrix}\right] \to (n', X', Y', Z', U')$$

$\hat{a}$, $\hat{b}$ - unbiased estimates of a, b.

$E\hat{a} = a, \quad E\hat{b} = b.$ will show later in a more general context

$\hat{f(x)}$ for a given x

$$\hat{f(x)} = \hat{a} + \hat{b}x$$

$$E\hat{f(x)} = E(\hat{a} + \hat{b}x) = E\hat{a} + E\hat{b}\cdot x =$$

$$= a + bx = f(x) \text{ - unbiased.}$$

Accuracy of $\hat{f(x)}$

$$E\left(\hat{f(x)} - \underbrace{f(x)}\right)^2 = Var\left(\hat{f(x)}\right) =$$

$$= \sigma^2\left[\frac{(x-\bar{x})^2}{U - \frac{X^2}{n}} + \frac{1}{n}\right] \qquad \bar{x} = \frac{X}{n}$$

$n \to \infty$?

$$U - \frac{X^2}{n} = (n-1)\,\overline{Var}(x) \approx n\cdot \overline{Var}(x)$$

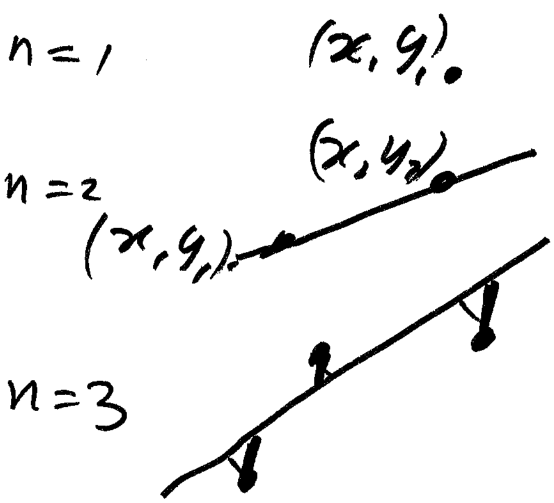
$$Var\left(\hat{f(x)}\right) \approx \frac{\sigma^2}{n}\left(\frac{(x-\bar{x})^2}{\overline{Var}(x)} + 1\right) \to 0$$

If σ^2 is <u>not known</u>.

$$Q(a,b) \sim \min_{a,b}$$

$$Q_{min} = \sum_i \left(y_i - (\hat{a} + \hat{b} x_i)\right)^2$$

$$E Q_{min} = \sigma^2 (n-2)$$ (Will show later)



can <u>not</u> est a, b

can est. a, b

can <u>not</u> est σ^2

can est a, b, σ^2

$$\hat{\sigma}^2 = \frac{Q_{min}}{n-2}$$

$$Q = \underline{n a^2} + 2 \underline{X a b} + \underline{U b^2}$$
$$- 2 \underline{Y a} - 2 \underline{Z b} + V$$

min. when: $na + Xb - Y = 0 \quad \times a$

$Xa + Ub - Z = 0 \quad \times b$

$$Q_{min} = V - Ya - Zb =$$

$$= V - Y \frac{Y - bX}{n} - Zb$$

$$= \left(V - \frac{Y^2}{n}\right) - \left(Z - \frac{XY}{n}\right) b$$

$$= \left(V - \frac{Y^2}{n}\right) - \frac{\left(Z - \frac{XY}{n}\right)^2}{U - \frac{X^2}{n}}$$

$$\hat{b} = \frac{Z - \frac{XY}{n}}{U - \frac{X^2}{n}}$$

$$\hat{a} = \frac{Y - bX}{n}$$

need to add

$$V = \sum_{i=1}^{n} y_i^2$$

and $\hat{\sigma}^2 = \frac{1}{n-2} Q_{min}$

$$\left.\begin{matrix}(x_1, y_1)\\ \vdots \\ (x_n, y_n)\end{matrix}\right] \longrightarrow (n, X, Y, Z, U, V) \quad \text{6 values}$$

$$\hat{b} = \frac{Z - \frac{XY}{n}}{U - \frac{X^2}{n}} \qquad \hat{a} = \frac{Y - \hat{b} X}{n}$$

for any given x: $\hat{f}(x) = \hat{a} + \hat{b} x$

$$\widehat{\mathrm{Var}}(\hat{f}(x)) =$$

$$= \underbrace{\frac{1}{n-2}\left[\left(V - \frac{Y^2}{n}\right) - \frac{\left(Z - \frac{XY}{n}\right)^2}{U - \frac{X^2}{n}}\right]}_{\hat{\sigma}^2}\left[\frac{\left(x - \frac{X}{n}\right)^2}{U - \frac{X^2}{n}} + \frac{1}{n}\right]$$

Curve fitting Problem.

$$y_i = f(x_i) + \varepsilon_i$$

$$f(x) = a_0 + a_1 x + \cdots a_k x^k$$

or

$$f(x) = b_0 + b_1 \cos x + b_2 \cos 2x + a_1 \sin x + a_2 \sin 2x$$

$$f_a(x) = a_1 f_1(x) + a_2 f_2(x) + \cdots a_m f_m(x)$$

$$a = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}$$

$$\underline{y_i = f_a(x_i) + \varepsilon_i} \qquad i = 1, \ldots n$$

$$f_a(x) = F(x) \cdot a$$

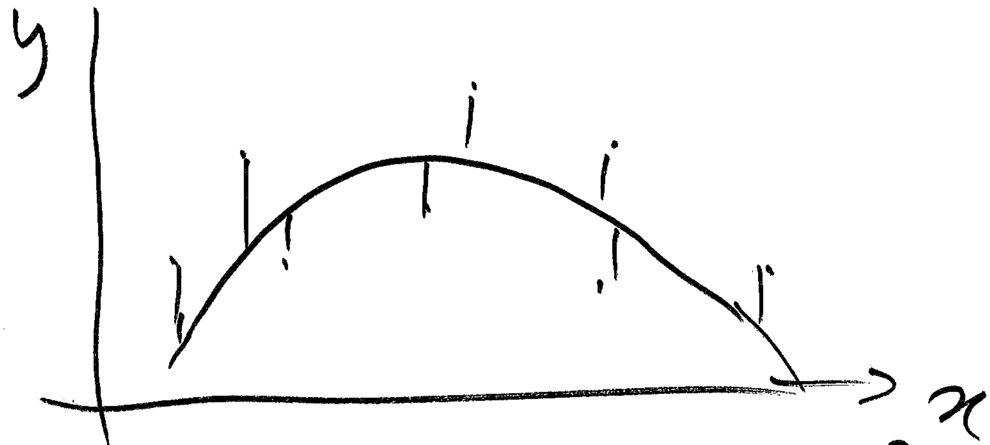
$$F(x) = \begin{bmatrix} f_1(x) & f_2(x) & \cdots & f_m(x) \end{bmatrix}$$

$y = B a + \varepsilon$ where

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \text{ - rand. vector} \qquad B = \begin{bmatrix} F(x_1) \\ F(x_2) \\ \vdots \\ F(x_n) \end{bmatrix} \ n \times m \text{ matrix}$$

$$B = \begin{bmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_m(x_1) \\ \vdots & & & \\ f_1(x_n) & f_2(x_n) & \cdots & f_m(x_n) \end{bmatrix} n \times m$$

ε_i - i.i.d. $\quad \varepsilon_i \sim (0, \sigma^2)$



$$Q(a) = \sum_i \left(y_i - f_a(x_i)\right)^2 \sim \min$$

y, z : inner product :

$$\langle y, z\rangle = \sum_{i=1}^{n} y_i z_i = y^T z \quad \Big| \quad y = Ba + \varepsilon$$

$$\|y\|^2 = \langle y, y\rangle$$

$$Q(a) = \|y - Ba\|^2 =$$

$$= \sum_{i=1}^{n} (y - Ba)_i^2 = \sum \left(y_i - (Ba)_i\right)^2$$

$$(Ba)_i = B_i a \qquad B = \begin{bmatrix} F(x_i) \end{bmatrix} \leftarrow i$$

i-th row of B: $B_i = F(x_i)$

$$Q(a) = \sum_{i=1}^{n} \Big(y_i - \underbrace{F(x_i)a}_{f_a(x_i)}\Big)^2 = \sum_{i=1}^{n} \big(y_i - f_a(x_i)\big)^2$$

$$Q(a) = \|y - Ba\|^2 =$$

$$= \langle y - Ba, y - Ba \rangle$$

$$= \langle y, y \rangle - 2\langle y, Ba \rangle + \langle Ba, Ba \rangle$$

$$= \|y\|^2 - 2\langle y, Ba \rangle + \|Ba\|^2$$

$\hat{a}$: $Q(\hat{a}) \sim \min$

Assume that $B^T B$ – invertible.

$$\|Ba - B(B^T B)^{-1} B^T y\|^2 =$$

$$= \|Ba\|^2 - 2\langle Ba, B(B^T B)^{-1} B^T y \rangle +$$

$$\|B(B^T B)^{-1} B^T y\|^2$$

$$\langle a, \underbrace{B^T B (B^T B)^{-1}}_{= I} B^T y \rangle = \langle Ba, y \rangle$$

$$\langle Ax, y \rangle = (Ax)^T y = x^T A^T y =$$

$$= \langle x, A^T y \rangle$$

$$\|Ba - B(B^TB)^{-1}B^Ty\|^2 =$$

$$\|Ba\|^2 - 2\langle Ba, y\rangle + \|B(B^TB)^{-1}B^Ty\|^2$$

$$\Rightarrow Q(a) = \|B(a - (B^TB)^{-1}B^Ty)\|^2$$
$$+ \|y\|^2 - \|B(B^TB)^{-1}B^Ty\|^2$$

$Q(a) \sim \min$ at $a = (B^TB)^{-1}B^Ty$

$$B = \begin{bmatrix} F(x_1) \\ \vdots \\ F(x_n) \end{bmatrix} = \begin{bmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_m(x_1) \\ \vdots & & & \\ f_1(x_n) & f_2(x_n) & \cdots & f_m(x_n) \end{bmatrix}$$

$f_1 \cdots f_m$ $\qquad \underline{n \times m}$ Large. $\qquad n \times m$

- linearly independent

When $n \geq m$ and x_i are choosen randomly then columns of B - independent.

$\Rightarrow$ rank $B = m$ $\Rightarrow$ ~~B^T~~
rank $B^T = m$ $\Rightarrow$ rank$(B^TB) = m$
B^TB - $m \times m$ matrix $\Rightarrow$ invertible

Since columns of B are indep $\Rightarrow$ $Ba = 0 \Rightarrow a = 0$

$$B = [B_1 \; B_2 \cdots B_m] \quad a = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}$$

$$Ba = a_1 B_1 + a_2 B_2 + \cdots + a_m B_m = 0$$

Since $B_1 \cdots B_m$ are indep $\Rightarrow$

$$a_1 = a_2 \cdots = a_m = 0$$

i.e. $a = 0$

$\Rightarrow$ $B^T B$ - invertible and

if $\|B(\;)\|^2 = 0 \Rightarrow (\;) = 0$

$$(\;) = a - (B^T B)^{-1} B^T y = 0$$

$\Rightarrow$ a is the only solution to $Q(a) \sim \min_a$

$$\boxed{a = (B^T B)^{-1} B^T y}$$