ML : MAXIMIZE $\ell(\mu) = \log p(x|\mu)$

BAYES (MAP): MAXIMIZE $p(\mu|x) = \dfrac{p(x|\mu)\,p(\mu)}{p(x)}$

EQUIVALENTLY MAXIMIZE:

$$f(\mu) \equiv \underbrace{\log p(\mu|X)}_{\text{LOG-POSTERIOR}} = \underbrace{\log p(x|\mu)}_{\text{LOG-LIKELIHOOD}} + \underbrace{\log p(\mu)}_{\text{LOG-PRIOR}} - C$$

NORMAL EXAMPLE CONT'D:

ASSUME $p(\mu) = N(\mu|0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\mu^2}$

$$\Longrightarrow \log p(\mu) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\mu^2$$

$$f(\mu) = \log p(x|\mu) + \log p(\mu)$$

$$= -\frac{1}{2}\sum_i (x_i-\mu)^2 - \frac{1}{2}\mu^2 \quad (+\text{CONSTANTS})$$

MAP-ESTIMATE $\hat{\mu}$:

$$\frac{d}{d\mu} f(\mu) = \sum_i x_i - n\mu - \mu = \sum_{i=1}^{n} x_i - (n+1)\mu = 0$$

$$\Longrightarrow \tilde{\mu} = \frac{1}{n+1}\sum_{i=1}^{n} x_i$$

DATA POINTS: $x_1, x_2, \ldots, x_m$ $\quad X = \{x_1, \ldots, x_m\}$

ASSUME $\quad x_i \overset{i.i.d.}{\sim} N(\mu, \underbrace{1}_{\text{VARIANCE } \sigma^2 \text{ IN GENERAL}}) \quad i=1,\ldots m$

PDF: $N(x_i \mid \mu, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i-\mu)^2}$ $\quad \Big\| \log\left(\frac{1}{\sqrt{2\pi}}\right) = \log(2\pi)^{-\frac{1}{2}} = -\frac{1}{2}\log(2\pi)$

LOG-PDF: $\log N(x_i \mid \mu, 1) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}(x_i-\mu)^2$

LIKELIHOOD FUNCTION

$$L(\mu) \equiv p(X \mid \mu) = N(x_1 \mid \mu, 1) N(x_2 \mid \mu, 1) \cdots N(x_m \mid \mu, 1) = \prod_{i=1}^{m} N(x_i \mid \mu, 1)$$

LOG-LIKELIHOOD

$$\ell(\mu) \equiv \log p(X \mid \mu) = \sum_{i=1}^{m} \log N(x_i \mid \mu, 1)$$

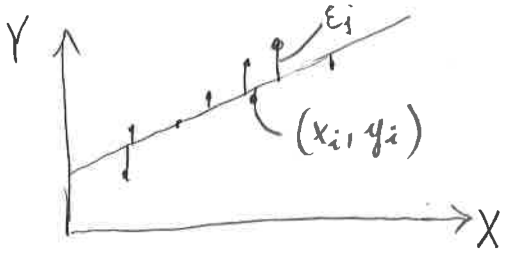
$$= \sum_{i=1}^{m} \left\{ -\frac{1}{2}\log(2\pi) - \frac{1}{2}(x_i-\mu)^2 \right\}$$

$$= \underbrace{-\frac{m}{2}\log(2\pi)}_{\substack{\text{CONSTANT, DOESN'T} \\ \text{DEPEND ON } \mu\text{, CAN BE DROPPED}}} - \frac{1}{2}\sum_{i=1}^{m}(x_i-\mu)^2$$

ML-ESTIMATE $\hat{\mu}$:

$$\frac{d}{d\mu}\ell(\mu) = -\frac{1}{2}\sum_{i=1}^{m} 2(x_i-\mu)(-1) = \sum_{i=1}^{m}\{x_i-\mu) = \sum_{i=1}^{m} x_i - m\mu = 0$$

$$\Rightarrow \hat{\mu} = \frac{1}{m}\sum_{i=1}^{m} x_i$$



DATA: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ OR $\{(x_i, y_i)\}_{i=1}^{n}$

MODEL: $y_i = a + bx_i + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2)$ || HERE $\Theta = (a, b)$

$y_i \sim N(a + bx_i, \sigma^2)$ || BECAUSE IF $Z \sim N(0, \sigma^2)$ THEN $Z + c \sim N(c, \sigma^2)$

$$p(y_i | a, b, x_i) = N(y_i | a + bx_i, \sigma^2)$$
$$= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y_i - a - bx_i)^2}$$

$$\log p(y_i | a, b, x_i) = -\frac{1}{2}\log(2\pi) - \log \sigma - \frac{1}{2\sigma^2}(y_i - a - bx_i)^2$$

LOG-LIKELIHOOD

DENOTE $X = (x_1, \ldots, x_n) \quad Y = (y_1, \ldots, y)$

$$l(a, b) = \log p(Y | a, b, X)$$
$$= \sum_{i=1}^{n} \log p(y_i | a, b, x_i)$$
$$= \sum_{i=1}^{n} \left\{ -\frac{1}{2}\log(2\pi) - \log \sigma - \frac{1}{2\sigma^2}(y_i - a - bx_i)^2 \right\}$$
$$= -\frac{n}{2}\log(2\pi) - n \log \sigma \underbrace{- \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - a - bx_i)^2}$$
$$= \ldots$$

MAXIMIZING THIS IS EQUIVALENT TO LS (LEAST SQUARES), WHERE $\sum_{i=1}^{n}(y_i - a - bx_i)^2$ IS MINIMIZED

ML-ESTIMATE

$$\frac{\partial}{\partial a} l(a, b) = 0$$

$$\frac{\partial}{\partial b} l(a, b) = 0$$