Squared cost error:

$$\boxed{E = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$$

Maximum likelihood:

- find μ where $X \sim (\mu, \sigma^2)$
- $$\boxed{p(x_i) = \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}}$$
- If all x_i $(x_1, x_2, \dots x_N)$ are i.i.d

$\rightarrow$ $$\boxed{p(x_1, x_2, \dots x_N) = \prod \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}}$$ joint prob.

- If we write this as likelihood:
  - Bayes':

$$\boxed{\underbrace{P(\mu|X)}_{\text{posteriori}} = \frac{\overbrace{P(X|\mu)}^{\text{likelihood}} \cdot P(\mu)}{P(X)}}$$

$\rightarrow$ evidence

$\hookrightarrow$ prior

- Likelihood:

$$p(X|\mu) = P(x_1, x_2, \dots x_N) = \prod \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$\rightarrow$ We want to find μ to maximize likelihood

$\rightarrow$ We want to find the best setting of μ so that the <u>data measured</u> is likely to come from this distribution.

How to solve? Use $\frac{\partial p}{\partial \mu} = 0$ or $\frac{\partial \log p}{\partial \mu} = 0$ (log likelihood)

<u>Steps</u>:

- log-likelihood:

$$\log p(X|\mu) = \log \prod \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$= \sum \left[\log \frac{1}{\sqrt{2\pi}\,\sigma} - \frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

$$= \sum \left[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(x_i-\mu)^2}{2\sigma^2}\right]$$ or $l = -\sum (x_i - \mu)^2$

$\hookrightarrow$ same form linear regression, just opposite sign $\rightarrow$ maximize.

- Take $\frac{\partial \log p}{\partial \mu} = 0$

$$\Leftrightarrow \sum \frac{(x_i - \mu)}{\sigma^2} = 0 \Leftrightarrow \boxed{\mu = \frac{\sum x_i}{N}}$$ $\rightarrow$ As expected

linear reg.: Minimize $E = \sum (y - \hat{y}_i)^2$

→ Moral of the story: minimize the squared error for linear regression = maximize the likelihood

- $y \sim \mathcal{N}(w^T x, \sigma^2)$

or $y = w^T x + \varepsilon$ where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$