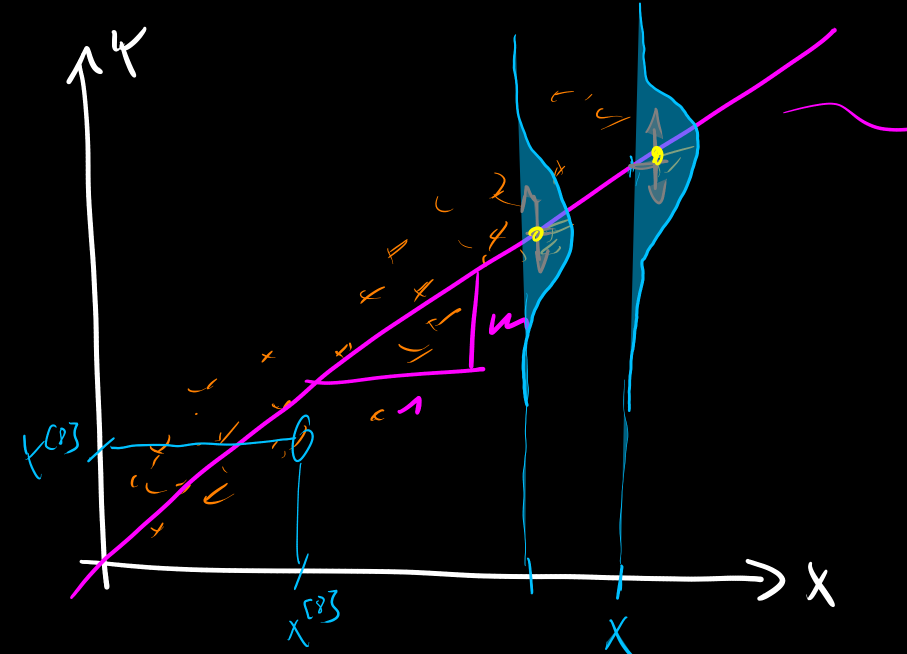
Linear Regression from a Probabilistic Perspective

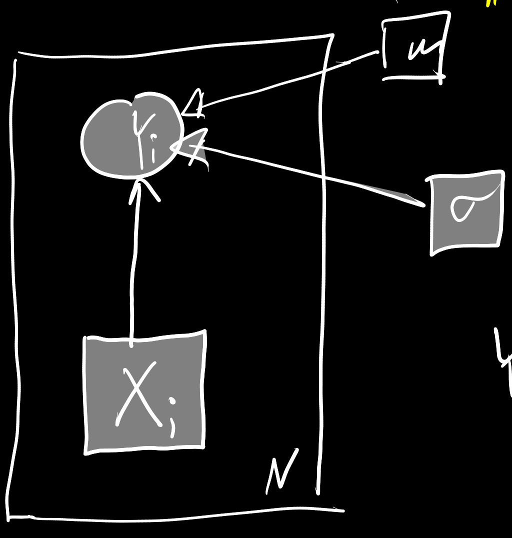


$Y \sim X$

$\Downarrow$ How?

$Y = m \cdot X$

slope of the linear function



"the spreading / distribution of Y given X is a Normal distribution"

$$Y_i \sim \mathcal{N}(Y_i, \mu = m \cdot X_i, \sigma)$$

$$p(Y_i \mid X_i, m, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(Y_i - \underbrace{m \cdot X_i}_{\mu})^2\right)$$

$$\mathcal{L}(D) \overset{\text{i.i.d.}}{=} \prod_{i=0}^{N-1} p(Y_i = y^{[i]} \mid X_i = x^{[i]}, m, \sigma)$$

$$D = \{\underbrace{(x^{[0]}, y^{[0]}), (x^{[1]}, y^{[1]}), \dots, (x^{[N-1]}, y^{[N-1]})}_{N \text{ samples}}\}$$

$$\mathcal{L}(D) = \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(y^{[i]} - m x^{[i]})^2\right)$$

$$= \left(\prod_{i=0}^{N-1}\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)\right) \cdot \left(\prod_{i=0}^{N-1} \exp\left(-\frac{1}{2\sigma}(y^{[i]} - m x^{[i]})^2\right)\right)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \cdot \exp\left(-\frac{1}{2\sigma}\sum_{i=0}^{N-1}(y^{[i]} - m x^{[i]})^2\right)$$

Log-Likelihood

$$\ell(D) = \log \mathcal{L}(D)$$

This was incorrect in the video

$$= -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma}\sum_{i=0}^{N-1}(y^{[i]} - m x^{[i]})^2$$

$$\stackrel{+}{=} -\frac{1}{2\sigma}\sum_{i=0}^{N-1}(y^{[i]} - m x^{[i]})^2$$

$$\sim -\frac{1}{2}\sum_{i=0}^{N-1}(y^{[i]} - m x^{[i]})^2$$

Goal: $m^* = \arg\max_m \ell(D)$

$$m^* = \arg\max_m \left(-\frac{1}{2}\sum_{i=0}^{N-1}(y^{[i]} - m x^{[i]})^2\right)$$

$$= \boxed{\arg\min_m \left(\underbrace{\frac{1}{2}\sum_{i=0}^{N-1}(y^{[i]} - m x^{[i]})^2}_{\hat{\ell}}\right)}$$

→ MLE under "a Gaussian error distribution"

This can be solved analytically ☺

$$\frac{\partial \hat{\ell}}{\partial m} = \cancel{\frac{1}{2}}\sum_{i=0}^{N-1}\cancel{2}(y^{[i]} - m x^{[i]}) \cdot (-x^{[i]}) \stackrel{!}{=} 0$$

$$= \sum_{i=0}^{N-1} -y^{[i]} x^{[i]} + m x^{[i]2} \stackrel{!}{=} 0$$

$$m \sum_{i=0}^{N-1} x^{[i]2} = \sum_{i=0}^{N-1} y^{[i]} x^{[i]}$$

$$\boxed{m^* = \frac{\sum_{i=0}^{N-1} y^{[i]} x^{[i]}}{\sum_{i=0}^{N-1} x^{[i]2}}}$$

→ for the univariate case, $x^{[i]} \in \mathbb{R}$

→ no regularization