Maximum Likelihood Estimation (MLE)

$$f_{Bayes} = \underset{f \in \{f \mid f : \mathcal{X} \to \mathcal{Y}\}}{\arg\min} L(f) \to = \mathbb{E}[\ell(f(\vec{X}), Y)]$$

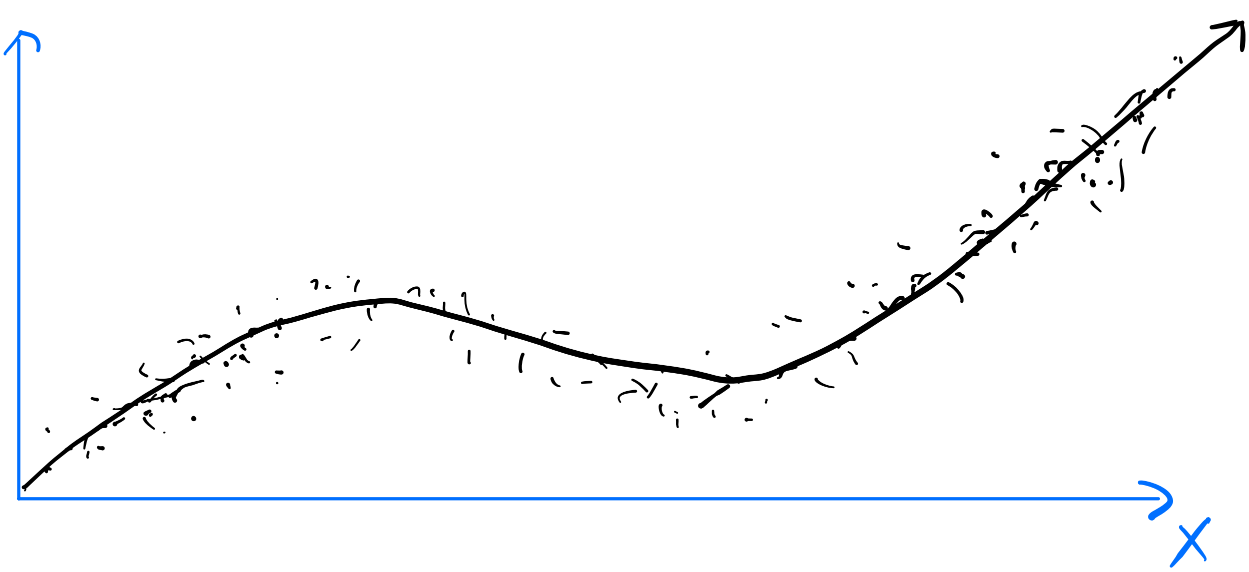
$\mathbb{P}_{\vec{X},Y}$ not known

assume squared loss and regression

$$f_{Bayes}(\vec{x}) = \mathbb{E}[Y \mid \vec{X} = \vec{x}]$$

$$= \int_{\mathcal{Y}} y \, P_{Y|\vec{X}=\vec{x}}(y \mid \vec{x}) \, dy$$

Lets use D to estimate $P_{Y|\vec{X}}$



MLE Basics

$D = (Z_1, \dots, Z_n) \in \mathcal{Z}^n$, $\mathbb{P}_D$, p_D

Z_i are i.i.d. with $\mathbb{P}_Z$ and pmf/pdf p_Z

Let $D = (z_1, \dots, z_n)$ be a fixed dataset

$$p_D(D) = p_D(z_1, \dots, z_n) \overset{\text{Independence}}{=} p_{Z_1}(z_1) \cdot p_{Z_2}(z_2) \cdot \dots \cdot p_{Z_n}(z_n)$$

$$\overset{\text{identical distribution}}{=} p_Z(z_1) \cdot p_Z(z_2) \cdot \dots \cdot p_Z(z_n)$$

$$= \prod_{i=1}^{n} p_Z(z_i)$$

$$p_{MLE} = \underset{p \in \mathcal{H}}{\operatorname{argmax}} \underbrace{\prod_{i=1}^{n} p(z_i)}_{\text{likelihood of } p \text{ give } D}$$

Ex: $Z_i \overset{i.i.d.}{\sim} \text{Bernoulli}(\alpha^*)$ $\qquad z_i \in \{0,1\}$

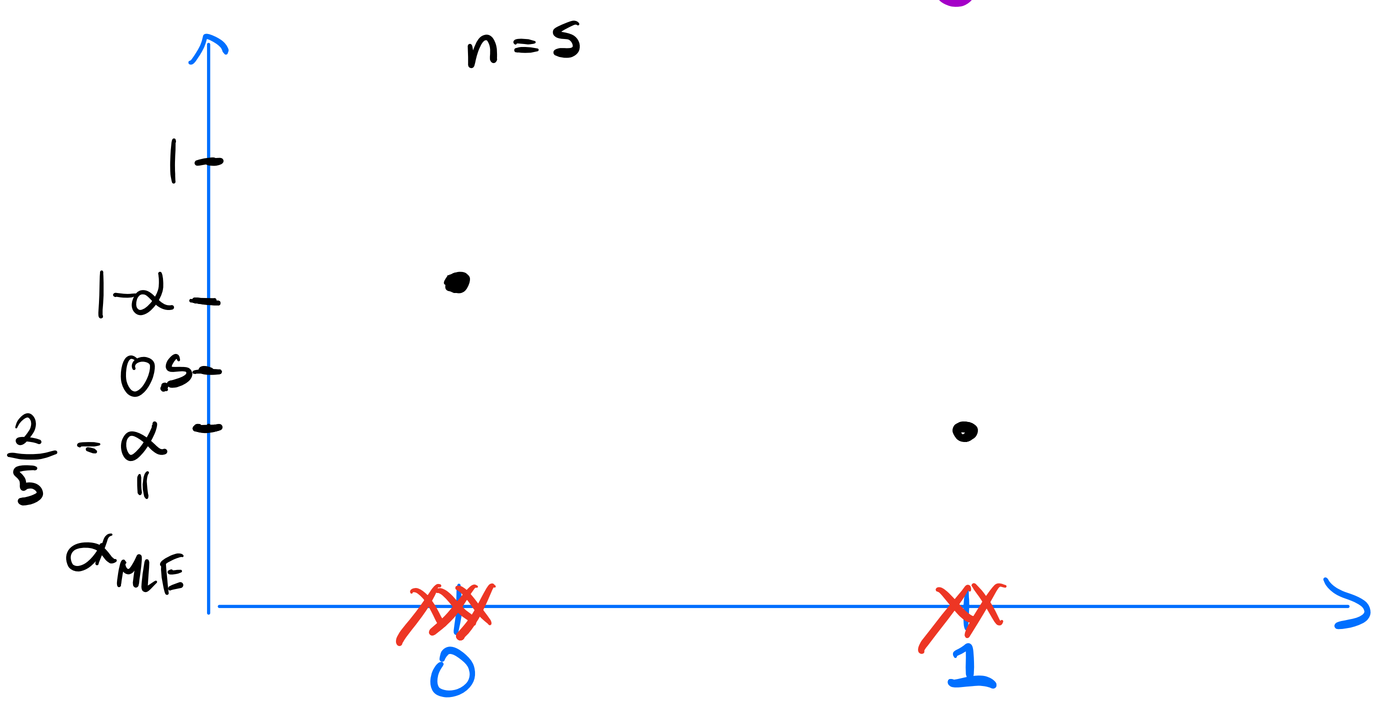
$$p_z(z) = (\alpha^*)^z (1-\alpha^*)^{(1-z)} = p_{\alpha^*}(z)$$

$$\mathcal{H} = \{p \mid p: \mathcal{Z} \to [0,1] \text{ and } p(z) = \alpha^z(1-\alpha)^{(1-z)},\ \alpha \in [0,1]\}$$

any $p_\alpha \in \mathcal{H}$ $\qquad p_\alpha(z) = \alpha^z(1-\alpha)^{(1-z)}$

$$p_{MLE} = p_{\alpha_{MLE}} \approx p_{\alpha^*} = p_z$$

$$\alpha_{MLE} = \underset{\alpha \in [0,1]}{\arg\max}\ \overbrace{\underbrace{\prod_{i=1}^{n} \alpha^{z_i}(1-\alpha)^{(1-z_i)}}_{= g(z)}}^{\text{likelihood } \alpha \text{ given } D}$$



Ex: $Z_i \sim \mathcal{N}(\mu^*, 1)$ is the height of the i-th person

$$p_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(z-\mu^*)^2}{2}\right) = p_{\mu^*}$$

$\exp(a) = e^a$

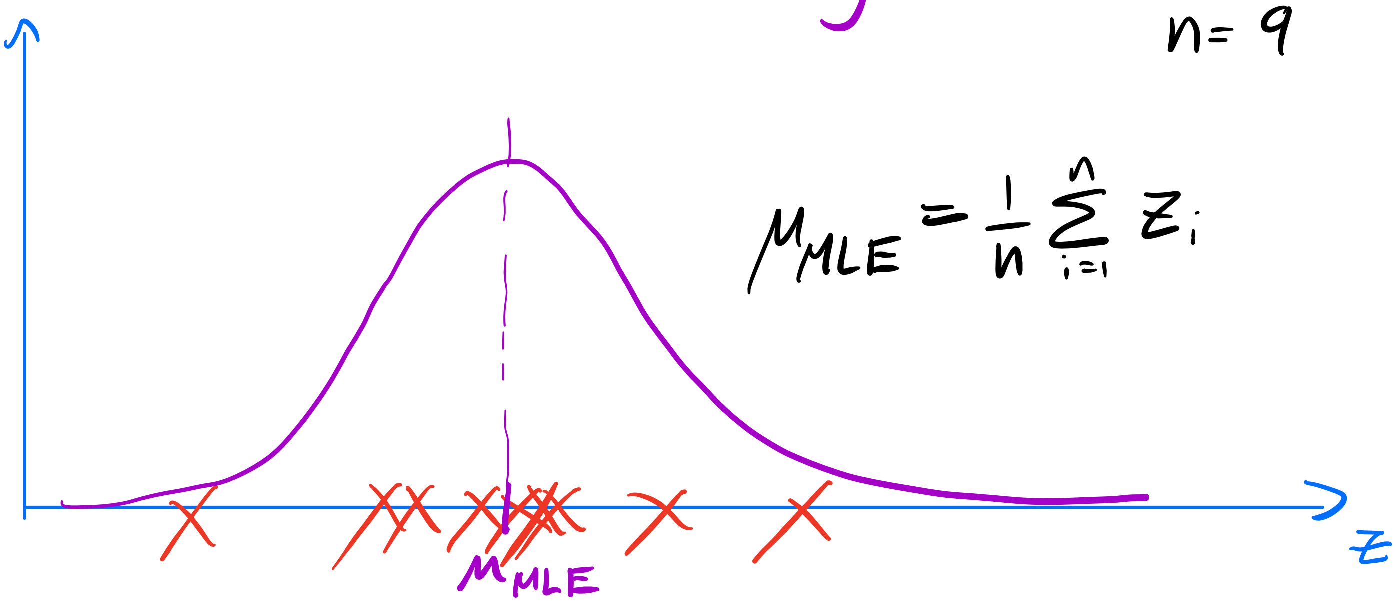
$$\mathcal{H} = \left\{ p \mid p : \mathcal{Z} \to [0,\infty) \text{ and } p(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(z-\mu^*)^2}{2}\right), \ \mu \in \mathbb{R} \right\}$$

$$p_{MLE} = p_{\mu_{MLE}} \approx p_{\mu^*} = p_Z$$

$$\mu_{MLE} = \operatorname*{argmax}_{\mu \in \mathbb{R}} \underbrace{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(z_i-\mu)^2}{2}\right)}_{= g(z)}$$

$n = 9$

$$\mu_{MLE} = \frac{1}{n} \sum_{i=1}^{n} z_i$$



$$\mu_{MLE} = \underset{\mu \in \mathbb{R}}{\arg\max} \prod_{i=1}^{n} p(z_i \mid \mu) \rightarrow = p_\mu(z_i)$$

$$= \underset{\mu \in \mathbb{R}}{\arg\max} \log\left(\prod_{i=1}^{n} p(z_i \mid \mu)\right)$$

monotone

$$= \underset{\mu \in \mathbb{R}}{\arg\max} \sum_{i=1}^{n} \log p(z_i \mid \mu)$$

$\log(a \cdot b) = \log(a) + \log(b)$

$$= \underset{\mu \in \mathbb{R}}{\arg\min} \underbrace{-\sum_{i=1}^{n} \log p(z_i \mid \mu)}_{\text{negative log likelihood}}$$

$$\underset{\mu \in \mathbb{R}}{\arg\max} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(z_i-\mu)^2}{2}\right)$$