Maximum A Posteriori (MAP) Estimate for Bernoulli Distribution

Weather (W) $\begin{cases} \to \text{good} \\ \to \text{bad} \end{cases}$

$W \in \{\underset{0}{\text{Good}}, \underset{1}{\text{Bad}}\}$ $\qquad W \sim \text{Bernoulli}(\theta)$

problem: we have data $D = \{G, G, B, G, B, B, \dots\}$

but no parameter θ

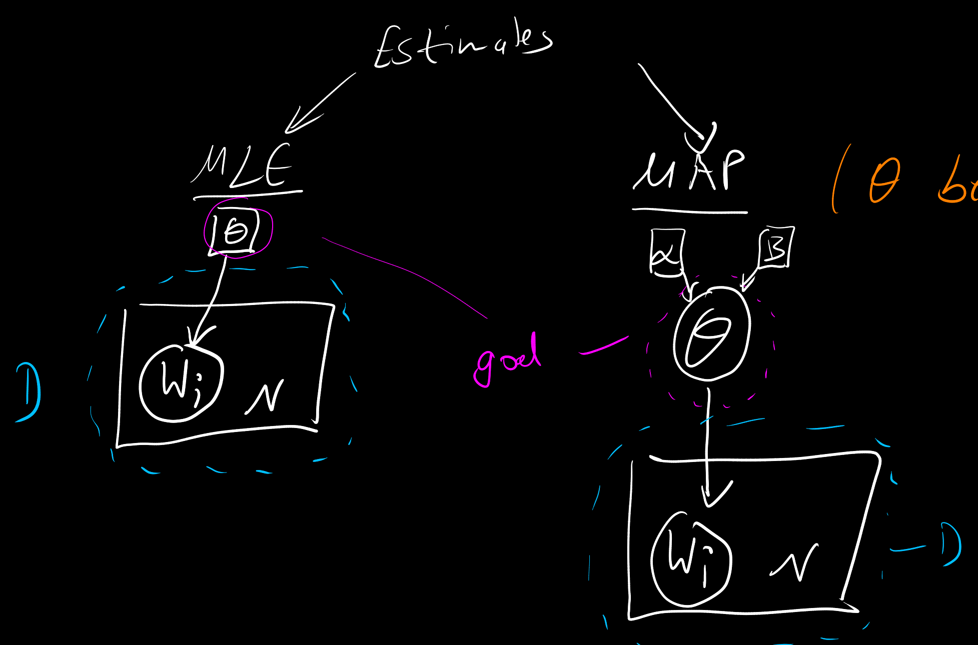
Task: infer θ from data D

$$\left[\ \text{MLE:} \quad \theta^* = \underset{\theta \in [0,1]}{\arg\max}\ \ell(D;\theta) \ \right]$$

can overfit!?

↳ not encoding any prior knowledge

Remedy: MAP



(θ becomes random too)

$\theta \in [0,1]$

$\theta \sim \text{Beta}(\alpha, \beta)$

joint distribution

$$p(\theta, \underbrace{\underline{W}}_{D}) = \underbrace{p(\underline{W} \mid \theta)}_{\text{likelihood}}\ \underbrace{p(\theta)}_{\text{prior}}$$

instead fit for parameter given data

$$\underset{\text{Bayes}}{p(\theta \mid D)} = \frac{p(D \mid \theta)\, p(\theta)}{p(D)}$$

hmm... difficult

$$\sim p(D \mid \theta)\, p(\theta)$$

$$p(\theta \mid D) \sim \left(\prod_{i=0}^{N-1} \text{Bern}(w^{[i]} \mid \theta) \right) \cdot \text{Beta}(\theta; \alpha, \beta)$$

$$= \left(\prod_{i=0}^{N-1} \theta^{w^{[i]}} \cdot (1-\theta)^{(1-w^{[i]})} \right) \cdot \frac{1}{\cancel{B(\alpha,\beta)}}\ \theta^{\alpha-1} \cdot (1-\theta)^{\beta-1}$$

$$\sim \left(\prod_{i=0}^{N-1} \theta^{w^{[i]}} \cdot (1-\theta)^{(1-w^{[i]})} \right) \cdot \theta^{\alpha-1} \cdot (1-\theta)^{\beta-1}$$

$$= M(D \mid \theta; \alpha, \beta)$$

$$\text{MAP} \quad \theta^*_{MAP} = \underset{\theta \in [0,1]}{\arg\max}\ M(D \mid \theta; \alpha, \beta)$$

take derivative and set to zero

$$m(D \mid \theta; \alpha, \beta) = \log(M(D \mid \theta; \alpha, \beta))$$

$$m(D \mid \theta; \alpha, \beta) = \log\left(\left(\prod_{i=0}^{N-1} \theta^{w^{[i]}} (1-\theta)^{(1-w^{[i]})} \right) \cdot \theta^{\alpha-1} \cdot (1-\theta)^{\beta-1} \right)$$

$$= \sum_{i=0}^{N-1} \left(\log(\theta^{w^{[i]}}) + \log\left((1-\theta)^{(1-w^{[i]})}\right) \right) + \log(\theta^{\alpha-1}) + \log\left((1-\theta)^{\beta-1}\right)$$

$$= \sum_{i=0}^{N-1} \left(w^{[i]} \log(\theta) + (1-w^{[i]}) \log(1-\theta) \right) + (\alpha-1)\log(\theta) + (\beta-1)\log(1-\theta)$$

$$\frac{\partial m}{\partial \theta} = \sum_{i=0}^{N-1} \left(\frac{w^{[i]}}{\theta} - \frac{1-w^{[i]}}{1-\theta} \right) + \frac{\alpha-1}{\theta} - \frac{\beta-1}{1-\theta} \overset{!}{=} 0$$

$$= \sum_{i=0}^{N-1} \left(\frac{w^{[i]}(1-\theta) - \theta(1-w^{[i]})}{\theta\cancel{(1-\theta)}} \right) + \frac{(\alpha-1)(1-\theta) - \theta(\beta-1)}{\theta\cancel{(1-\theta)}} \overset{!}{=} 0$$

$$= \sum_{i=0}^{N-1} \left(w^{[i]} - \cancel{w^{[i]}\theta} - \theta + \cancel{\theta w^{[i]}} \right) + \alpha - \alpha\theta - 1 + \theta - \theta\beta + \theta$$

$$= \sum_{i=0}^{N-1} (w^{[i]} - \theta) + \alpha - \alpha\theta - \beta\theta - 1 + 2\theta \overset{!}{=} 0$$

$$= \sum_{i=0}^{N-1} w^{[i]} - \underbrace{\sum_{i=0}^{N-1} \theta}_{N\theta} + \alpha - \alpha\theta - \beta\theta - 1 + 2\theta$$

$$= \sum_{i=0}^{N-1} w^{[i]} + \alpha - 1 - \theta(N + \alpha + \beta - 2) \overset{!}{=} 0$$

$$\theta^*_{MAP} = \frac{\sum_{i=0}^{N-1} w^{[i]} + \alpha - 1}{N + \alpha + \beta - 2}$$