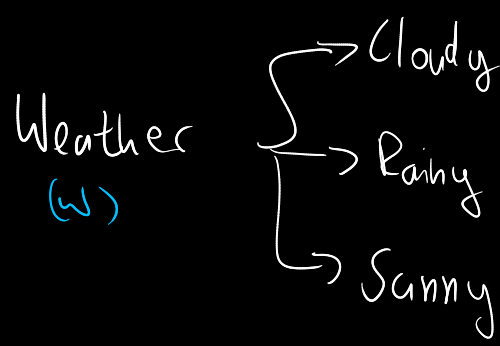


MLE for the categorical



$$w \in \{ \overset{0}{C}, \overset{1}{R}, \overset{2}{S} \}$$

$$p(w) = \prod_{d=0}^{D-1} \theta_d^{I(w=d)}$$

?

given observations

$$D = \{ 0, 0, 1, 1, 2, 0, 2, 2, 1, 0, \dots \}$$

what is θ ?

Likelihood

$$\begin{aligned} \mathcal{L}(D; \theta) &\stackrel{i.i.d.}{=} \prod_{i=0}^{N-1} p(w=w^{[i]}) \\ &= \prod_{i=0}^{N-1} \prod_{d=0}^{D-1} \theta_d^{I(w^{[i]}=d)} \end{aligned}$$

Log-Likelihood

$$\begin{aligned} \ell(D; \theta) &= \log(\mathcal{L}(D; \theta)) \\ &= \sum_{i=0}^{N-1} \sum_{d=0}^{D-1} I(w^{[i]}=d) \log(\theta_d) \end{aligned}$$

Maximum Likelihood Estimate (MLE)

$$\theta^* = \underset{\theta \in [0,1]^D}{\operatorname{argmax}} (\ell(D; \theta))$$

constrained optimization

$$\sum_{d=0}^{D-1} \theta_d = 1$$

→ incorporate constraints into target function

→ Lagrange Multiplier

constraint function

$$\begin{aligned} \sum_{d=0}^{D-1} \theta_d - 1 &= 0 \\ g(\theta) &= 1 - \sum_{d=0}^{D-1} \theta_d \end{aligned}$$

Augmented target

$$\begin{aligned} \hat{\mathcal{L}}(D; \theta, \lambda) &= \ell(D; \theta) + \lambda g(\theta) \\ &= \sum_{i=0}^{N-1} \sum_{d=0}^{D-1} (I(w^{[i]}=d) \log(\theta_d)) + \lambda \left(1 - \sum_{d=0}^{D-1} \theta_d \right) \end{aligned}$$

→ Take derivative and set to 0

$$\frac{\partial \hat{\mathcal{L}}}{\partial \theta_e} = \sum_{i=0}^{N-1} \left(I(w^{[i]}=e) \frac{1}{\theta_e} \right) - \lambda \stackrel{!}{=} 0$$

θ_e ... one component of θ

$$\Leftrightarrow \lambda = \frac{1}{\theta_e} \sum_{i=0}^{N-1} I(w^{[i]}=e)$$

N_e ... the # of times the weather is 'e'

$$\lambda = \frac{N_e}{\theta_e}$$

$$\frac{\partial \hat{\mathcal{L}}}{\partial \lambda} = 1 - \sum_{d=0}^{D-1} \theta_d$$

How to get rid of the λ ?

$$\Leftrightarrow 1 = \sum_{d=0}^{D-1} \theta_d$$

some trick:

$$\lambda = \lambda - \lambda \cdot 1 = \lambda \cdot \sum_{d=0}^{D-1} \theta_d = \sum_{d=0}^{D-1} \lambda \theta_d = \sum_{d=0}^{D-1} \frac{N_e}{\theta_e} \theta_d$$

$$\lambda = \frac{N_e}{\theta_e} = \text{const.} = \frac{N_{e-1}}{\theta_{e-1}} = \frac{N_{e+1}}{\theta_{e+1}} = \left(\frac{N_d}{\theta_d} \right)$$

$$\lambda = \sum_{d=0}^{D-1} \frac{N_d}{\theta_d} \cdot \theta_d = \sum_{d=0}^{D-1} N_d = \boxed{N}$$

$$N = \frac{N_e}{\theta_e} \Leftrightarrow \boxed{\theta_e = \frac{N_e}{N}}$$