Esempio. $\mathcal{X} = \{$viola, bianco, giallo$\}$

$\mathcal{Y} = \{$Asteracee, Orchidacee, Rubiacee$\}$

100 osservaz:

- 40 sono Asteracee di cui 5 fiori viola, 20 bianchi e 15 gialli
- 40 sono Orchidacee di cui 20 viole, 10 bianchi e 10 gialli
- 20 sono Rubiacee di cui tutte con fiori gialli

(D)

$y \backslash x$ — $p(x,y)$	viola	bianco	giallo
$\pi_A = 40\%$ A	5% (1/20)	20% (1/5)	15% (3/20)
$\pi_O = 40\%$ O	20% (1/5)	10% (1/10)	10% (1/10)
$\pi_R = 20\%$ R	0%	0%	20% (1/5)
$p(x)$	25% (1/4)	30% (3/10)	45% (9/20)

$p(x,y) \rightarrow$ (1/20)

$p(x) \rightarrow$ (1/4)

$y \backslash x$ — $p(y\|x)$	V	b	g
A	1/5	2/3	1/3
O	4/5	1/3	2/9
R	0	0	4/9

$$p(y|x) = \frac{p(x,y)}{p(x)}$$

$h^*(\text{viola}) = 0$ $\quad$ $Pr[h^*(x) \neq y \mid x = \text{viola}] = 1/5$

$h^*(\text{bianco}) = A$ $\quad$ $Pr[h^*(x) \neq y \mid x = \text{bianco}] = 1/3$

$h^*(\text{giallo}) = R$ $\quad$ $Pr[h^*(x) \neq y \mid x = \text{giallo}] = 5/9$

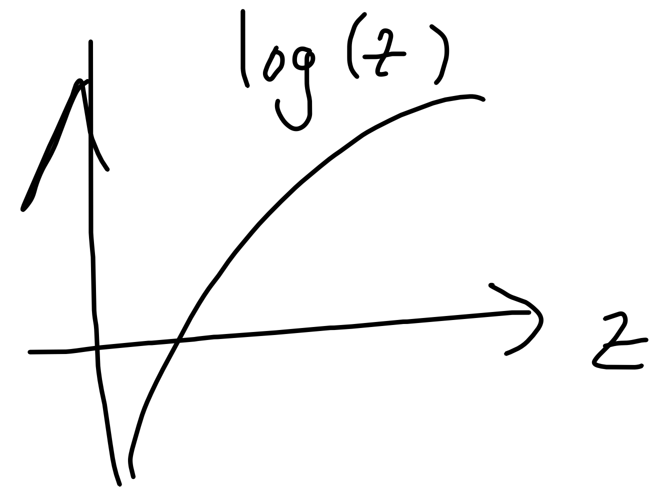
Quanto vale $L_D(h^*)$?

(inaccuratezza classificatore)

$$Pr[h^*(x) \neq y] = 1/5 \cdot \underbrace{1/4} + 1/3 \cdot 3/10 + 5/9 \cdot 9/20 = 8/20 = 40\%$$

$$\underset{j \in \mathcal{Y}}{\operatorname{argmax}} \frac{\pi_j P_j(x)}{\underbrace{p(x)}_{\text{non dipende da } j}} = \underset{j \in \mathcal{Y}}{\operatorname{argmax}} \; \pi_j P_j(x)$$

$$= \underset{j}{\operatorname{argmax}} \; \log(\pi_j P_j(x))$$



Maximum Likelihood Estimation (MLE)

$$\mathcal{L}(\mu,\sigma) = \prod_{i=1}^{m} p(x^{(i)};\mu,\sigma) =$$

$x^{(1)},\ldots,x^{(m)}$

$$= \prod_{i=1}^{m} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{(x^{(i)}-\mu)^2}{2\sigma^2}\right)$$

$$\underset{\mu,\sigma}{\operatorname{argmax}}\ \mathcal{L}(\mu,\sigma) = \underset{\mu,\sigma}{\operatorname{argmax}}\ \log\mathcal{L}(\mu,\sigma)$$

$$= \underset{\mu,\sigma}{\operatorname{argmin}} \sum_{i=1}^{m}\left[\underbrace{+\frac{1}{2}\log(2\pi\sigma^2)}_{\text{non dipende da } \mu} + \frac{(x^{(i)}-\mu)^2}{2\sigma^2}\right]$$

Immaginiamo <u>σ fissato</u>

$$\operatorname*{argmin}_{\mu} \sum_{i=1}^{m} \left[\frac{1}{2} \log(2\pi\sigma^2) + \frac{(x^{(i)}-\mu)^2}{2\sigma^2} \right]$$

$$= \operatorname*{argmin}_{\mu} \cancel{\frac{1}{2\sigma^2}} \underbrace{\sum_{i=1}^{m} (x^{(i)}-\mu)^2}_{g(\mu) \text{ (convessa)}} \qquad \nabla g(\mu) \overset{?}{=} 0$$

$$\nabla g(\mu) = g'(\mu) = \cancel{2} \sum_{i=1}^{m} (x^{(i)}-\mu)(-1) \overset{?}{=} 0$$

$$\Longleftrightarrow \sum_{i=1}^{m} x^{(i)} = \sum_{i=1}^{m} \mu = m \cdot \mu \Rightarrow \hat{\mu} = \frac{1}{m} \sum_{i=1}^{m} x^{(i)}$$

$$\arg\min_{\sigma} \underbrace{\frac{1}{\sigma^2}\sum_{i=1}^{m} (x^{(i)}-\hat{\mu})^2 + \underbrace{\sum_{i=1}^{m} 2\log(2\pi\sigma)}_{2m\log(2\pi\sigma)}}_{G(\sigma)}$$

$$\nabla G(\sigma) \overset{?}{=} 0 \iff -\cancel{2}\sigma^{-3}\sum_{i=1}^{m}(x^{(i)}-\hat{\mu})^2 + \cancel{2}m\frac{1}{\cancel{2\pi}\sigma}\cdot\cancel{2\pi} = 0$$

$$\sigma^{-3}\sum_{i=1}^{m}(x^{(i)}-\hat{\mu})^2 = \sigma^{-1} m$$

$$\iff \frac{1}{m}\sum_{i=1}^{m}(x^{(i)}-\hat{\mu})^2 = \sigma^2 = \hat{\sigma}^2$$