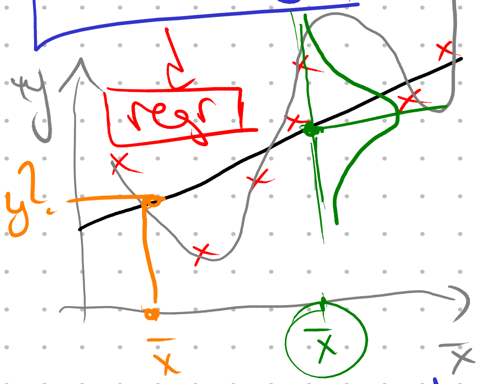
ML

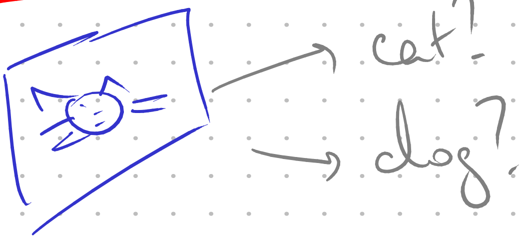
Supervised learning

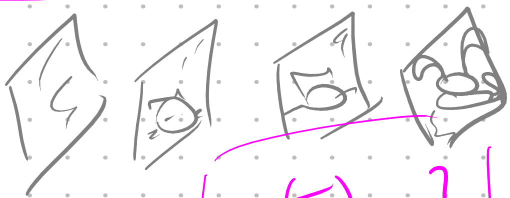
Reinforcement learning

Unsupervised learning



Classification





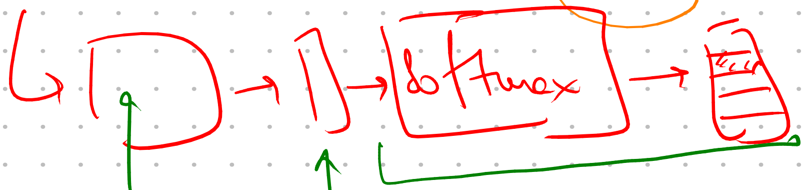
$p(\bar{x}) = ?$

clustering

$D = \{(\bar{x}_n, y_n)\}_{n=1}^{N}$

$f: \bar{x} \mapsto y$

$p(y|\bar{x}) = ?$



feature extraction

features

$$p(x=a, y=b)$$

$$p(x) = \int p(x,y)\, dy$$ ← marginalization

conditional independence

$$p(x,y|z) = p(x|z)\,p(y|z)$$

$$p(x,y) = p(x)\,p(y)$$

independence

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

$$p(x,y) = p(x|y)\,p(y) = p(y|x)\,p(x)$$

$$\boxed{p(x|y) = \frac{p(y|x)\,p(x)}{p(y)}}$$ ← Bayes Theorem

Пар-ры модели данные

$$\underbrace{p(\theta|D)}_{\text{posterior distribution}} = \frac{\overbrace{p(D|\theta)}^{\text{likelihood}}\,\overbrace{p(\theta)}^{\text{prior}}}{p(D)}$$

$$\underbrace{p(D)}_{\text{evidence}} = \int p(D|\theta)\,p(\theta)\,d\theta$$

$$p(\theta|D) \propto p(D|\theta)\,p(\theta)$$

(1) ML – max likelihood

$$\theta_{ML} = \arg\max_{\theta} p(D|\theta)$$

(2) MAP – maximum a posteriori

$$\theta_{MAP} = \arg\max_{\theta} p(\theta|D) =$$

$$= \arg\max_{\theta} p(D|\theta)\,p(\theta)$$

(3) Predictive distribution

Bayesian inference

$$p(x|D) = \int p(x,\theta|D)\,d\theta =$$

$$= \int p(x|\theta,\underline{\underline{D}}) \cdot \underbrace{p(\theta|D)}_{\text{posterior}}\,d\theta =$$

$$= \int \underbrace{p(x|\theta)}_{\text{likelihood}}\;\underbrace{p(\theta|D)}_{\text{posterior}}\,d\underline{\underline{\theta}}$$

$$p(D|\theta) = \prod_{d\in D} p(d|\theta)$$

$\mathbb{E}_{p(x)}[f(x)] = \int f(x)\cdot p(x)dx$

$$p(x|D) = \mathbb{E}_{p(\theta|D)}\left[p(x|\theta)\right] \approx$$

$$\approx \frac{1}{R}\sum_{r=1}^{R} p(x|\theta^{(r)}),$$ где $\theta^{(r)} \sim p(\theta|D)$

Монетка

$\theta = p(\text{орёл})$

$1-\theta = p(\text{решка})$

$D = h\,t\,t\,h\,h\,t\,t\,t$

$$p(D|\theta) = \prod_{d\in D} p(d|\theta) = \theta^n (1-\theta)^m,$$ где

$n = \#$ орлов

$m = \#$ решек

$$\theta_{ML} = \arg\max_\theta p(D|\theta) \stackrel{?}{=} \frac{n}{n+m}$$

$$\frac{\partial p(D|\theta)}{\partial\theta} = n\theta^{n-1}(1-\theta)^m - m\theta^n(1-\theta)^{m-1} = 0$$

$$\theta^{n-1}(1-\theta)^{m-1}\left(n(1-\theta) - m\theta\right) = 0$$

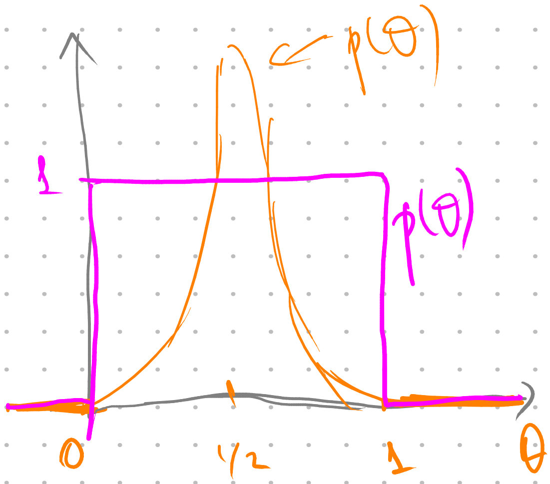
$$\boxed{\theta_{ML} = \frac{n}{n+m}}$$

$$\theta = 0,\ 1,\ \frac{n}{n+m}$$

$D = h$

$p(D|\theta) = \theta \longrightarrow \max$

$\theta_{ML} = 1$



prior

$p(\theta)$

$$p(\theta) = \begin{cases} 1, & \theta \in [0,1] \\ 0, & \theta \notin [0,1] \end{cases}$$

$$p(\theta|D) \underline{\underline{\propto}} p(D|\theta)\underline{\underline{p(\theta)}} = \begin{cases} \theta^n (1-\theta)^m, & \theta \in [0,1] \\ \underline{\underline{0}}, & \theta \notin [0,1] \end{cases}$$

$$p(D) = \int_{-\infty}^{\infty} p(D|\theta)\, p(\theta)\, d\theta = \int_0^1 \theta^n (1-\theta)^m \, d\theta = B(n+1, m+1)$$

$$= \frac{\Gamma(n+1)\,\Gamma(m+1)}{\Gamma(n+m+2)} = \frac{n!\, m!}{(n+m+1)!}$$

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$$p(\theta|D) = \begin{cases} \frac{(n+m+1)!}{n!\, m!} \theta^n (1-\theta)^m, & \theta \in [0,1] \\ 0, & \theta \notin [0,1] \end{cases}$$

$$\theta_{MAP} = \frac{n}{n+m}$$

$b = h$

$$(3)\quad p(\text{open}\mid D) = \int \underline{p(\text{open}\mid\theta)}\, p(\theta\mid D)\, d\theta =$$

$$= \int_0^1 \theta \cdot \frac{(n+m+1)!}{n!\,m!}\,\theta^n (1-\theta)^m \, d\theta =$$

$$= \frac{(n+m+1)!}{n!\,m!} \int_0^1 \theta^{n+1} (1-\theta)^m \, d\theta =$$

$$= \frac{\cancel{(n+m+1)}!}{\cancel{n!}\,\cancel{m!}} \cdot \frac{(n+1)\cancel{!}\cdot \cancel{m!}}{(n+m+2)\cancel{!}}$$

$$\boxed{p(\text{open}\mid D) = \frac{n+1}{n+m+2}}$$

Laplace's rule