Всем привет! ~ UData School 2018

Вот так

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

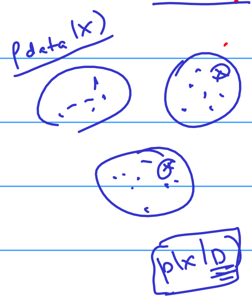
posterior

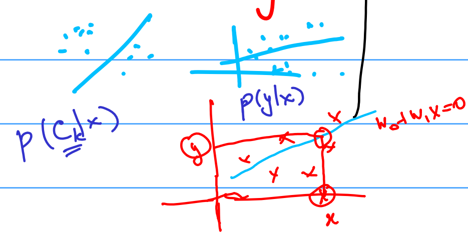
$$p(D)=\int p(D|\theta)p(\theta)d\theta$$

AI → ..., ML

ML → Unsup., Supervised

Supervised → Class., Regr.





ML: $p(D|\theta)\xrightarrow[\theta]{}\max$ θ_{ML}

MAP: $p(\theta|D)\xrightarrow[\theta]{}\max$ θ_{MAP}

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

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

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

$$\underbrace{p(x|D)}_{\text{predictive distribution}}=\int p(x|\theta)p(\theta|D)\,d\theta\propto\int p(x|\theta)p(\theta|D)p(\theta)\,d\theta$$

θ – „вероятность орла" heads tails

$\boxed{p(D|\theta)}$

$D=t$

$p(t|\theta)=\theta$ $\quad p(h|\theta)=1-\theta$

$p(ttht|\theta)=\theta^3(1-\theta)$

Bernoulli trials

$$p(D|\theta)=\theta^{n_t}(1-\theta)^{n_h}$$

ML: $\boxed{\theta_{ML}=\frac{n_t}{n_h+n_t}}$

$$\theta_{ML}=\arg\max_\theta p(D|\theta)=\arg\max_\theta \theta^{n_t}(1-\theta)^{n_h}$$

$$n_t\cdot\theta^{n_t-1}(1-\theta)^{n_h}-n_h(1-\theta)^{n_h-1}\theta^{n_t}=0$$

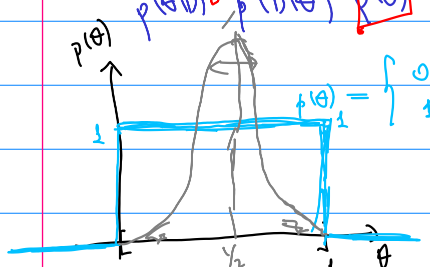
$$\theta^{n_t-1}(1-\theta)^{n_h-1}\left(n_t\cdot(1-\theta)-n_h\cdot\theta\right)=0$$

$p(D|\theta)=\theta^2\to\max$

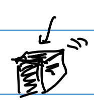
$D=tt$: $\theta_{ML}=1$

$\theta=0$ $\quad\theta=1$ $\quad\boxed{\theta_{ML}=\frac{n_t}{n_h+n_t}}$

$$p(\theta|D)\propto p(D|\theta)\cdot\overbrace{\boxed{p(\theta)}}^{\text{prior}}$$



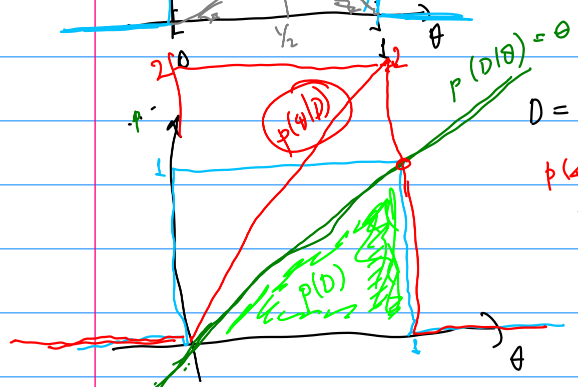
$$p(\theta)=\begin{cases}0, & \theta\notin[0,1]\\ 1, & \theta\in[0,1]\end{cases}$$



$$p(D|\theta)=\theta^{n_t}(1-\theta)^{n_h}\quad\text{- likelihood}$$

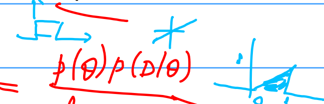
$$p(\theta|D)\propto\begin{cases}\theta^{n_t}(1-\theta)^{n_h}, & \theta\in[0,1]\\ 0, & \theta\notin[0,1]\end{cases}$$

$$\theta_{MAP}=\arg\max_\theta p(\theta|D)=\frac{n_t}{n_h+n_t}$$



$D=$ „t"

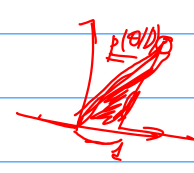
$$p(\theta|D)=\frac{p(\theta)p(D|\theta)}{\int p(\theta)p(D|\theta)d\theta}$$

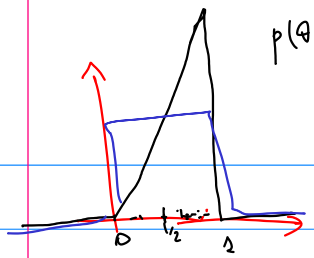


$p(\theta|D)\propto\theta,\ [0,1]$

$$p(\theta|D)=\begin{cases}2\theta, & \theta\in[0,1]\\ 0, & \theta\notin[0,1]\end{cases}$$

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





$p(\theta|D)$

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

$D = n_h \ldots,\ n_t \ldots$

$$p(D|\theta) = \theta^{n_t}(1-\theta)^{n_h}$$

$$p(\theta|D) = \begin{cases} 0, & \theta \notin [0,1] \\ \dfrac{\theta^{n_t}(1-\theta)^{n_h}}{p(D)}, & \theta \in [0,1] \end{cases}$$

θ_{ML} $\quad \theta_{MAP}$

$p(\theta) =$ 

$$p(D) = \int_0^1 \theta^{n_t}(1-\theta)^{n_h}\, d\theta$$

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}$$

$$\Gamma(n) = (n-1)!$$

$$p(D) = \int_0^1 \theta^{n_t}(1-\theta)^{n_h}\, d\theta = \frac{n_t!\, n_h!}{(n_t+n_h+1)!}$$

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

$$p(t|D) = \int p(t,\theta|D)\, d\theta = \int_{-\infty}^{\infty} \underbrace{p(t|\theta, D)}_{\theta}\, p(\theta|D)\, d\theta =$$

$$= \int_0^1 \theta \cdot \frac{(n_t+n_h+1)!}{n_t!\, n_h!} \cdot \theta^{n_t}(1-\theta)^{n_h}\, d\theta = \ldots \int_0^1 \theta^{n_t+1}(1-\theta)^{n_h}\, d\theta = \frac{(n_t+n_h+1)!}{n_t!\, n_h!} \cdot \frac{(n_t+1)!\, n_h!}{(n_t+n_h+2)!} = \boxed{\frac{n_t+1}{n_t+n_h+2}}$$

правило Лапласа

$\theta_{ML}, \theta_{MAP} \rightarrow p(t|D)$

$$\frac{n_t}{n_h+n_t} \qquad \frac{n_t+1}{n_h+n_t+2}$$

Bayesian smoothing

 k граней $\bar\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ $\quad \sum \theta_i = 1,\ \theta_i \geq 0$

$$p(D|\bar\theta) = \theta_1^{n_1}\theta_2^{n_2}\ldots\theta_k^{n_k} \qquad p(\bar\theta) = 1,\ \bar\theta \in \ldots$$

$D = \{n_1, n_2, \ldots, n_k\}$

$$p(i|D) = \boxed{\frac{n_i+1}{\sum n_j + k}}$$

Language models

$p(w \mid w_1, \ldots, w_d)$

n-grams

#{всем_пробел}

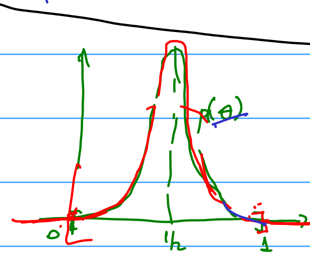
#{пробел_колонн}

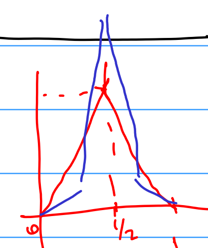
$p(w|\ldots)$

$p([- - - w_1 w_2 - - -] \mid \bar\theta) = 0$

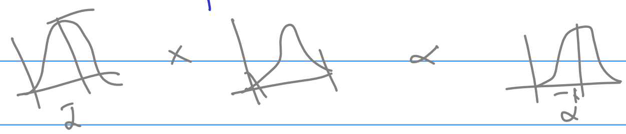
$$p(D|\theta) = \theta^{n_t}(1-\theta)^{n_h} \qquad p(\theta|D),\ \theta_{MAP},\ p(t|D)$$

$\boxed{p(\theta) = \{1, 0 : \ldots\}}$





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



$$p(\theta|\alpha) \cdot \cancel{p(D|\theta)} \propto \underbrace{p(\theta|D) = p(\theta|\alpha')}\ \cdot\ p(D'|\theta) \propto \underbrace{p(\theta|\alpha'')}$$

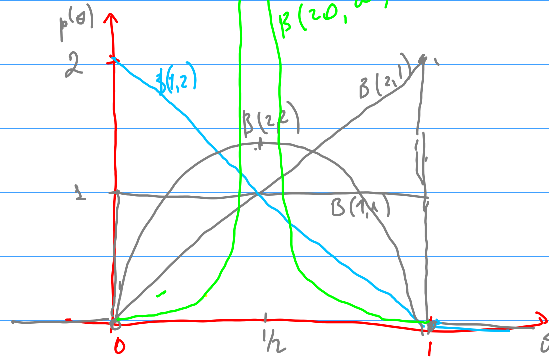
Conjugate priors: семейство $p(\theta|\alpha)$ наз. сопр. для $p(D|\theta)$, если $p(\theta|\alpha)\cdot p(D|\theta) \propto p(\theta|\alpha')$

$$\underline{p(\theta)} \cdot \underbrace{\theta^{n_t}\cdot(1-\theta)^{n_h}}_{p(D|\theta)} \approx p'(\theta)$$

$$\boxed{\theta^{\alpha-1}(1-\theta)^{\beta-1}} \cdot \theta^{n_t}(1-\theta)^{n_h} = \theta^{\overset{\alpha'}{\alpha+n_t-1}} \cdot (1-\theta)^{\overset{\beta'}{\beta+n_h-1}}$$

beta distribution

$$p(\theta) = \begin{cases} \dfrac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}, & \theta \in [0,1] \\ 0, & \theta \notin [0,1] \end{cases}$$



$\vec{D}$

$B(1,1)$, $p(\theta)=1$ $\rightsquigarrow D=\{n_h, n_t\} \rightsquigarrow p(\theta|D)=B(n_t+1, n_h+1)$

$p(D|\theta)=\theta^{n_t}(1-\theta)^{n_h}$

$p(\theta; 1,1)$

$p(\theta; 2,1)$

$p(\theta; 1,2)$

$p(\theta; 2,2) \propto \theta(1-\theta)$

$$\theta^{\alpha-1}(1-\theta)^{\beta-1} \xrightarrow[\theta]{} \max$$

$$\theta^* = \frac{\alpha-1}{\alpha+\beta-2}$$

$$p(\theta) = B(20,20)$$

$$B(x; \tfrac{1}{2}, \tfrac{1}{2}) \propto x^{-\frac{1}{2}}(1-x)^{-\frac{1}{2}} = \frac{1}{\sqrt{x(1-x)}}$$

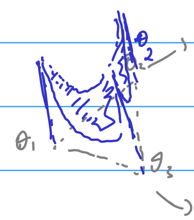
$$p(D|\bar\theta) = \theta_1^{n_1}\theta_2^{n_2}\dots\theta_k^{n_k}$$

$$p(\bar\theta) \propto \theta_1^{\alpha_1-1}\theta_2^{\alpha_2-1}\dots\theta_k^{\alpha_k-1}$$

Dirichlet $\quad p(\bar\theta) = \dfrac{\text{---}}{Dir(\bar\alpha)}$

$$Dir(\alpha_1 \dots \alpha_k) = \frac{\Gamma(\alpha_1)\cdot\Gamma(\alpha_2)\dots\Gamma(\alpha_k)}{\Gamma(\alpha_1+\dots+\alpha_k)}$$

$$p(\bar\theta; \bar\alpha)\cdot p(D|\bar\theta) \propto p(\bar\theta; \bar\alpha+\bar n)$$



d1000

topic modeling

$$Dir\left(\tfrac{1}{10^3}, \tfrac{1}{10^3}, \tfrac{1}{10^3}, \dots, \tfrac{1}{10^3}\right)$$