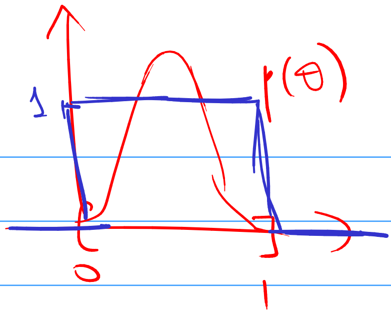
$p(\bar{x}|D) = ?$

likelihood, prior

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



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

$$p(D|\theta) = \theta^{n}(1-\theta)^{m} \qquad \theta_{ML} = \frac{n}{n+m}$$

пар. мод, данните

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

$$\boxed{p(D|\underline{\theta}) = \prod_{n} p(x_n|\underline{\theta})}$$

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

predictive distribution

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

$$= \int p(\bar{x}|\bar{\theta},D)\cdot p(\bar{\theta}|D)\,d\bar{\theta}$$

$$= \int p(\bar{x}|\bar{\theta})\,p(\bar{\theta}|D)\,d\bar{\theta} = \mathbb{E}_{p(\bar{\theta}|D)}\left[p(\bar{x}|\bar{\theta})\right]$$

$$= \int p(\bar{x}|\bar{\theta})\cdot\frac{p(D|\bar{\theta})\,p(\bar{\theta})}{p(D)}\,d\bar{\theta}$$

$$\boxed{\Gamma(a+1) = a!}$$

$$\propto \int p(\bar{x}|\bar{\theta})\,p(\bar{\theta})\,p(D|\bar{\theta})\,d\bar{\theta}$$

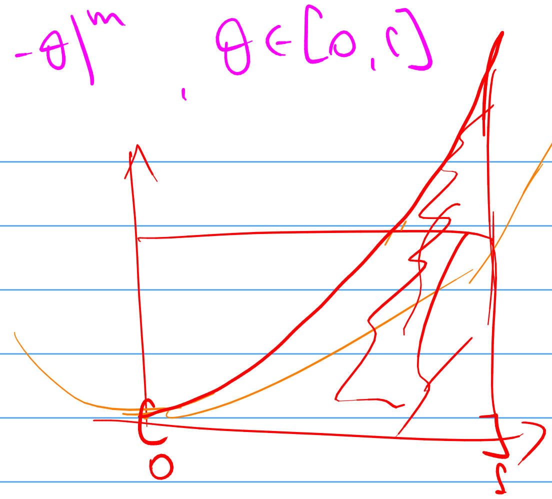
$$p(D) = \int_{-\infty}^{\infty} p(D|\theta)\,p(\theta)\,d\theta = \int_{0}^{1}\theta^{n}(1-\theta)^{m}\,d\theta = B(n+1,m+1)$$

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

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

~~Бета-распределение~~

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



$$p(\text{heads}|D) =$$

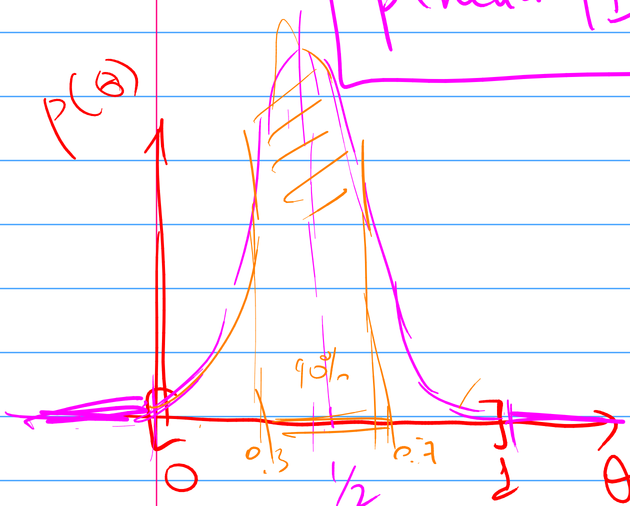
$$= \int p(\text{heads}|\theta)\, p(\theta|D)\, d\theta =$$

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

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

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

Laplace's Rule



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

Conjugate priors

$$p(\theta|\bar{\alpha})$$

$$p(\theta|\alpha') = p(\theta|D) \propto p(D|\theta)\,p(\theta|\alpha)$$

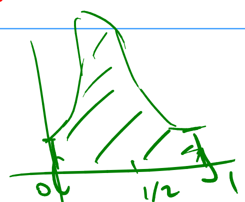
$(\alpha+n, \beta+m)$ hthh (α,β)

$$p(\theta|D) \qquad p(D|\theta) \qquad p(\theta)$$

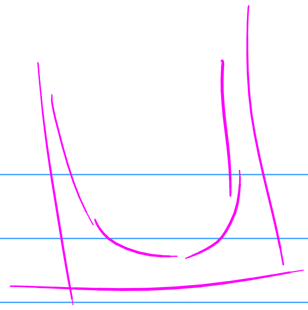
$$\theta^{\alpha+n-1}(1-\theta)^{\beta+m-1} \propto \theta^{n}(1-\theta)^{m} \times \frac{1}{B(\alpha,\beta)}\,\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$B(\alpha+n, \beta+m)$

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



$$p(x \mid \alpha, \beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$



$$\theta_k = (1 - \sum \theta_i)$$

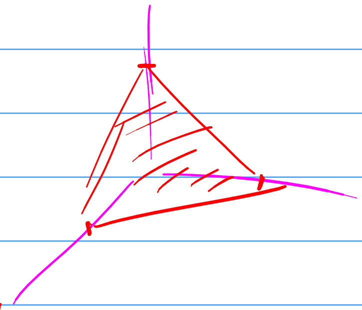
$$\bar\theta = (\theta_1, \dots, \theta_k)$$

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

Dirichlet distribution

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

$$\forall i\ \theta_i \geq 0 \qquad \sum \theta_i = 1$$



$$p\left(\bar\theta \;\middle|\; \left(\frac{1}{100}, \dots, \frac{1}{100}\right)\right)$$

$k = 500$

Sparsity

$\boxed{\text{LDA}}$ – latent Dirichlet allocation

$$X_1, X_2, \dots, X_n \qquad p \qquad D$$

$$\hat p_k(D) = \frac{\#\{\text{heads после } k \text{ подряд}\}}{\#\{k \text{ подряд}\}} = \frac{\sum_{t \in I_k(D)} X_t}{|I_k(D)|}$$

$$I_k(D) = \{ t \mid t \text{ после } k \text{ подряд} \}$$

$$p \stackrel{?}{=} E\left[\hat p_k(D) \mid I_k(D) \neq \emptyset\right] = ?$$

$t, \tau \in 1..n$

$$E[\hat p \mid I_k \neq \emptyset] = p(X_\tau = 1 \mid I_k(D) \neq \emptyset) =$$

$$\tau \sim \mathrm{Unif}(I_k(D))$$

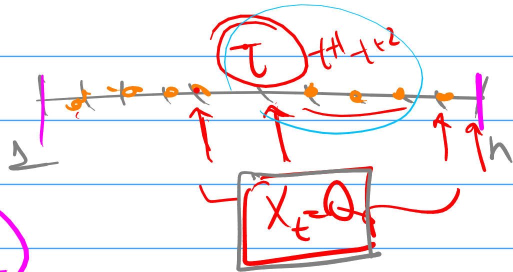
$$= p\left(X_t = 1 \;\middle|\; t = \tau, \prod_{i=t-k}^{t-1} X_i = 1\right) \propto$$

$$\propto p(t=\tau \mid X_t=1, \Pi_{..}=1)\; p(X_t=1 \mid \Pi_{..}=1)$$

p

$$p(X_\tau=1 \mid I_k\neq\emptyset) \propto p\cdot p(t=\tau \mid X_t=1, \Pi_{..}=1)$$

$$p(X_\tau=0 \mid I_k\neq\emptyset) \propto (1-p)\cdot p(t=\tau \mid X_t=0, \Pi_{..}=1)$$



$$p(t=\tau \mid X_t=0, \Pi_{..}=1) >$$

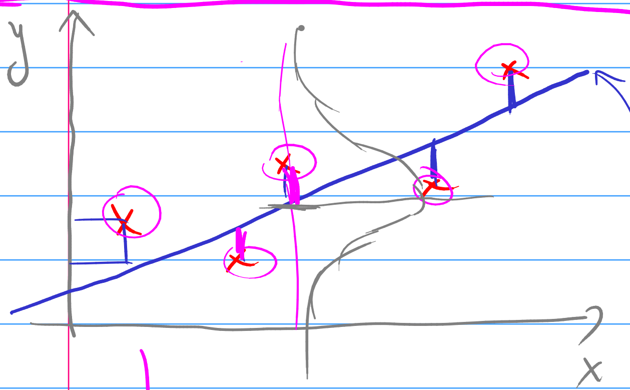
$$> p(t=\tau \mid X_t=1, \Pi_{..}=1)$$

$k=1$, $N=3$

6 12

	$\hat{p}$
hhh	1
hht	1/2
hth	0
thh	1
htt	0
tht	0
tth	
ttt	

$$\frac{5/2}{6} = \frac{5}{12}$$



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

$$\hat{y} = w_0 + w_1 x$$

$$\bar{x} \mapsto \begin{pmatrix} \bar{x} \\ 1 \end{pmatrix}$$

$$\hat{y} = \bar{x}^T \bar{w}$$

$$y = \bar{w}^T \bar{x} + \varepsilon$$

$$\boxed{\varepsilon \sim N(0, \sigma^2)}$$

$$L = \sum_{n=1}^{N} (y_n - \hat{y}_n)^2 = \sum_{n=1}^{N} (y_n - \bar{w}^T \bar{x}_n)^2 \xrightarrow[\bar{w}]{} \min$$

$$L = \begin{pmatrix} y_1 - \bar{x}_1^T \bar{w} \\ \vdots \\ y_N - \bar{x}_N^T \bar{w} \end{pmatrix}^T \begin{pmatrix} \\ \vdots \\ \\ \end{pmatrix}$$

$$\bar{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} - \overset{= X}{\begin{pmatrix} x_{11} & \cdots & x_{1d} \\ & \ddots & \\ x_{N1} & \cdots & x_{Nd} \end{pmatrix}} \overset{= \bar{w}}{\begin{pmatrix} w_1 \\ \vdots \\ w_d \end{pmatrix}}$$

$N \times 1 \qquad N \times d \qquad d \times 1$

$$L = (\bar{y} - X\bar{w})^T (\bar{y} - X\bar{w}) =$$

$$= \bar{y}^T \bar{y} - 2\underbrace{\bar{w}^T X^T \bar{y}}_{=} - \underbrace{\cancel{\bar{y}^T X \bar{w}}}_{= (\bar{y}^T X \bar{w})^T} + \bar{w}^T (X^T X) \bar{w} \xrightarrow[\bar{w}]{} \min$$

$$\nabla_{\bar{w}} f = \begin{pmatrix} \partial f / \partial w_1 \\ \vdots \\ \partial f / \partial w_d \end{pmatrix}$$

$$\nabla_{\bar{w}} (\underbrace{\bar{w}^T \bar{a}}_{= \sum a_i w_i}) = \bar{a}$$

$$\nabla_{\bar{w}} (\bar{w}^T \bar{w}) = 2\bar{w}$$

$$\nabla_{\bar{w}} (\bar{w}^T A \bar{w}) = A\bar{w} + A^T \bar{w} = (A + A^T)\bar{w}$$

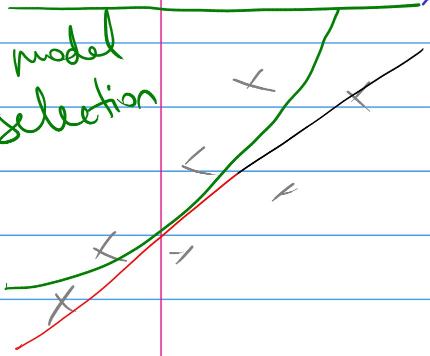
$$\nabla_{\bar{w}} (\underline{\bar{w}^T A \bar{w}}) = \begin{pmatrix} \vdots \\ \bar{a}_{*k}^T \bar{w} \\ \vdots \end{pmatrix} + \begin{pmatrix} \vdots \\ \bar{a}_{k*}^T \bar{w} \\ \vdots \end{pmatrix}$$

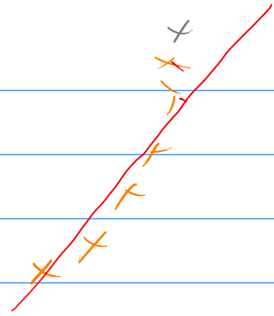
$$\frac{\partial \left(\sum_{i,j} a_{ij} w_i w_j \right)}{\partial w_k} = \frac{\partial \left(\sum_{i \neq k} a_{ik} w_i w_k + \sum_{j \neq k} a_{kj} w_k w_j + a_{kk} w_k^2 \right)}{\partial w_k}$$

$$= \sum_{i \neq k} a_{ik} w_i + \sum_{j \neq k} a_{kj} w_j + 2 a_{kk} w_k = \boxed{\sum_i a_{ik} w_i} + \sum_j a_{kj} w_j$$

$$\nabla_{\bar w} L(\bar w) = -2X^T\bar y + 2(X^TX)\bar w^* = 0$$

model selection



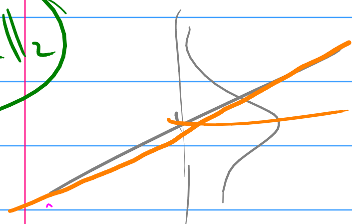


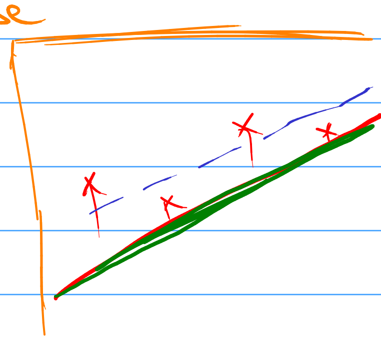
$$(X^TX)\bar w^* = X^T\bar y$$

$$\boxed{\bar w^* = (X^TX)^{-1}X^T\bar y}$$

Moore-Penrose pseudoinverse

$\|\bar x - \hat x\|_2$





$$p(D|\bar w) = \prod_{n=1}^{N} p(y_n|\bar w, \bar x_n) =$$

$$= \prod_{n=1}^{N} \mathcal{N}(y_n|\bar w^T\bar x_n, \sigma^2) =$$

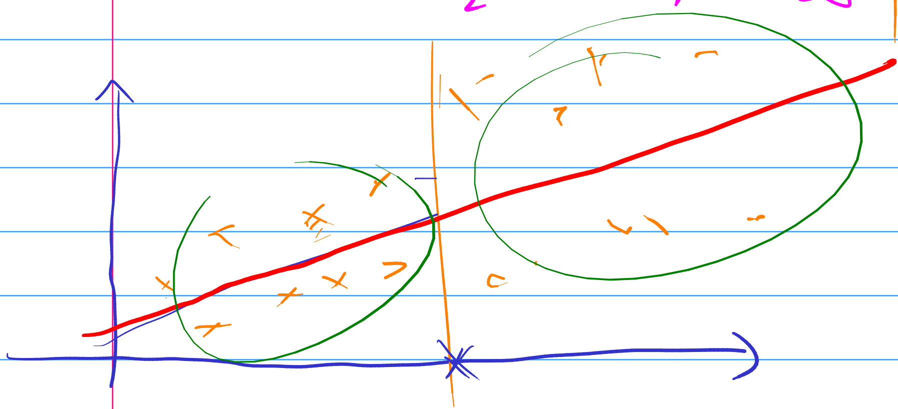
$$= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2\sigma_n^2}(y_n-\bar w^T\bar x_n)^2}$$

$$\ln p(D|\bar w) = -\frac{n}{2}\ln(2\pi\sigma^2) - \sum_{n=1}^{N}\frac{1}{2\sigma_n^2}(y_n-\bar w^T\bar x_n)^2$$

$\xleftarrow{\bar w}$ max

$$= -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\boxed{\sum_{n=1}^{N}(y_n-\bar w^T\bar x_n)^2}$$

$\xrightarrow[\bar w]{}$ min



$$\sum_{n=1}^{N}\frac{1}{\sigma_n^2}(y_n-\bar w^T\bar x_n)^2$$