posterior, prior, likelihood

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

$$p(D\mid\theta) = \theta^n (1-\theta)^m$$

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

predictive distribution

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

$$p(\theta \mid D) \propto \begin{cases} \theta^n (1-\theta)^m, & \theta \in [0,1] \\ 0, & \text{—} \end{cases}$$

$$p(x\mid\theta) = \frac{n+1}{n+m+2}$$

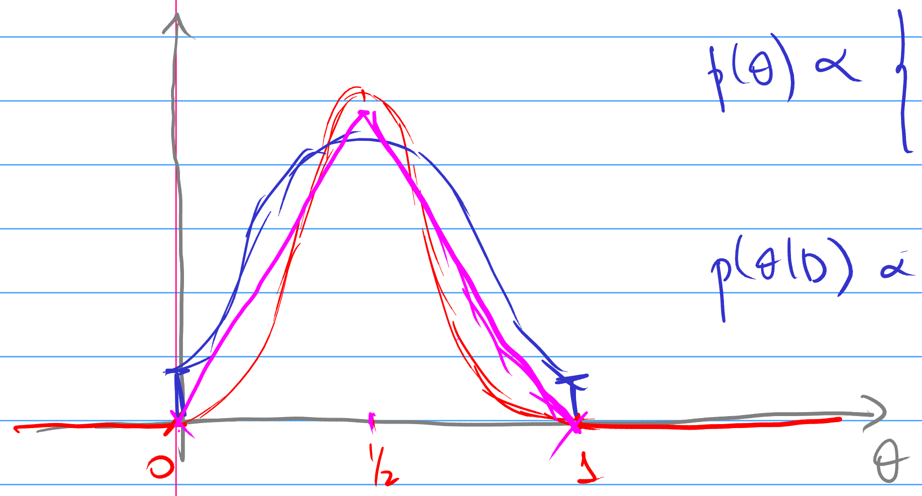
$$p(d \mid t) = \frac{\overset{0.95}{p(t\mid d)}\;\overset{0.01}{p(d)}}{p(t\mid d)p(d) + \underbrace{p(t\mid \bar d)}_{0.05} p(\bar d)} \approx 1/6$$

false positives:

$$p(t \mid \bar d) = 0.05$$

false negatives:

$$p(\bar t \mid d) = 0.05$$



$$p(\theta) \propto \begin{cases} e^{-\frac{1}{2\sigma^2}(\theta - \frac12)^2}, & \theta \in [0,1] \\ 0, & \theta \notin [0,1] \end{cases}$$

$$p(\theta \mid D) \propto p(\theta)\, p(D\mid\theta) =$$

$$= e^{-c\cdot(\theta-\frac12)^2} \cdot \theta^n (1-\theta)^m$$

$$p(\theta) = \begin{cases} 2\theta, & \theta \in [0, \frac12] \\ 2-2\theta, & \theta \in [\frac12, 1] \\ 0, & \text{—} \end{cases}$$

$$p(\theta \mid D) \propto \begin{cases} \theta^{n+1}(1-\theta)^m \\ \theta^n (1-\theta)^{m+1} \\ 0, & \text{—} \end{cases}$$

$$p(\theta) \propto \theta(1-\theta)$$

сем-во сопряж. априорн. распр.

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

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

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

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

$p(h \mid \theta) = \theta, \; p(t \mid \theta) = 1-\theta$

$$\underbrace{\mathrm{Beta}(\theta \mid \alpha, \beta)}_{p(\theta)} \times \underbrace{\theta^{n}(1-\theta)^{m}}_{p(D \mid \theta)} \propto \underbrace{\mathrm{Beta}(\theta \mid \alpha+n, \beta+m)}_{\boxed{p(\theta \mid D)}}$$

$$p(\theta \mid D) \times \underbrace{\theta^{n'}(1-\theta)^{m'}}_{p(D' \mid \theta)} \propto \underbrace{\mathrm{Beta}(\theta \mid \alpha+n+n', \beta+m+m')}_{p(\theta \mid D, D')}$$

$$\mathrm{Beta}(\theta \mid \tfrac{1}{2}, \tfrac{1}{2}) \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}} = \frac{1}{\sqrt{\theta(1-\theta)}}$$

K yonlü

$\bar{\theta} = (\theta_1, \ldots, \theta_K)$

$$p(D \mid \bar{\theta}) = \theta_1^{n_1}\theta_2^{n_2}\ldots\theta_K^{n_K}$$

$$\times\; p(\bar{\theta} \mid \bar{\alpha}) = \mathrm{Dir}(\bar{\theta} \mid \bar{\alpha}) = \frac{1}{\mathrm{Dir}(\bar{\alpha})}\theta_1^{\alpha_1-1}\theta_2^{\alpha_2-1}\ldots\theta_K^{\alpha_K-1}$$

$$p(\bar{\theta} \mid D) = \mathrm{Dir}(\bar{\theta} \mid \alpha_1+n_1, \alpha_2+n_2, \ldots, \alpha_K+n_K)$$

$$\mathrm{Dir}(\bar{\theta} \mid \underbrace{\tfrac{1}{10}, \ldots, \tfrac{1}{10}}_{K})$$

$K = 30000$

$\boxed{N=3}$

$\boxed{k=1}$

k=3

hhh	1
hht	1/2
hth	0
htt	0
thh	1
tht	0
~~tth~~	
~~ttt~~	

$$\text{avg} = \frac{5}{12}$$

$D = \ldots\ldots$

$$\mathbb{E}\left[\hat{p}_k(D) \mid I_k(D) \neq \emptyset\right]$$

$$\hat{p}_k(D) = \frac{\#[h \text{ none } \overset{k}{h\ldots h}]}{\#[\text{beg. none } \overset{k}{h\ldots h}]}$$

$$\text{"} \; I_k(D) = \{\text{beg} \mid \text{none } \overset{k}{h\ldots h}\}$$

1) $\mathbb{E}\left[\hat{p}_k(D) \mid I_k(D) = \emptyset\right] = p(X_\tau = 1 \mid I_k(D) \neq \emptyset)$ ye

$\mid \tau \sim \text{Unif}(I_k(D))$

$\text{"}\{i \mid \prod x_j = 1\}$

2) $p(X_\tau = 1 \mid I_k(D) \neq \emptyset) = p(X_t = 1 \mid \tau = t, I_k(D) \neq \emptyset) =$

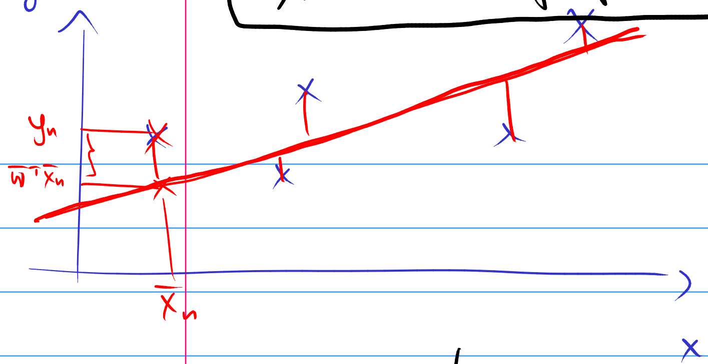
$$= p\left(X_t = 1 \mid \tau = t, \prod_{i=t-k}^{t-1} X_i = 1\right) \propto$$

$$\propto p\left(\tau = t \mid X_t = 1, \prod_{i=t-k}^{t-1} X_i = 1\right) p\left(X_t = 1 \mid \prod_{t-k}^{t-1} X_i = 1\right) \quad \text{"} p$$

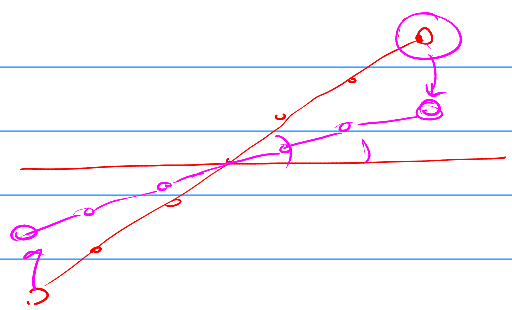
$$p(X_\tau = 1 \mid I_k(D) \neq \emptyset) \propto p \cdot p\left(\tau = t \mid X_t = 1, \prod X_i = 1\right)$$

$$p(X_\tau = 0 \mid \ldots) \propto (1-p) \cdot p\left(\tau = t \mid X_t = 0, \prod X_i = 1\right)$$

Линейная регрессия



Regression to the mean



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

$$\boxed{y_n} \approx w_1 x_{n1} + w_2 x_{n2} + \ldots + w_d x_{nd} = \bar{w}^T \bar{x}_n$$

$$\sum_{n=1}^{N} |y_n - \bar{w}^T \bar{x}_n| \longrightarrow \min$$

$$\boxed{\sum_{n=1}^{N} (y_n - \bar{w}^T \bar{x}_n)^2 \longrightarrow \min \qquad \text{МНК}}$$

$$\sum_{n=1}^{N} \ln |y_n - \bar{w}^T \bar{x}_n| \longrightarrow \min$$

МНК:

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

$$\begin{pmatrix} \ldots \\ y_n - \bar{x}_n^T \bar{w} \\ \ldots \end{pmatrix}^T \begin{pmatrix} \ldots \\ y_n - \bar{x}_n^T \bar{w} \\ \ldots \end{pmatrix}$$

$$\bar{y} - \begin{pmatrix} — \bar{x}_1 — \\ — \bar{x}_n — \\ — \bar{x}_N — \end{pmatrix} \bar{w}$$

$$(\bar{y} - X\bar{w})^T (\bar{y} - X\bar{w}) \xrightarrow[\bar{w}]{} \min$$

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

$$\nabla_{\bar{w}} L(\bar{w}) = ?$$

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

$$\bar{w}^T \bar{a} = w_1 a_1 + \ldots + w_d a_d$$

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

$$\bar{w}^T \bar{w} = w_1^2 + \ldots + w_d^2$$

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

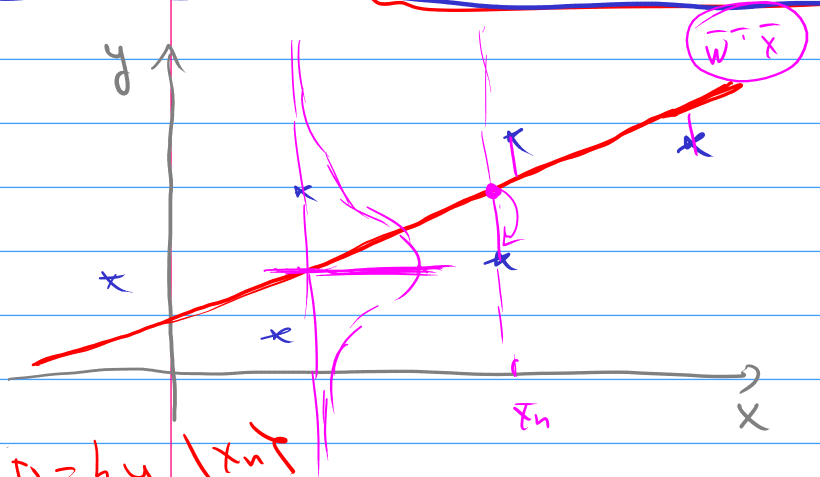
$$\sum_{i,j} a_{ij} w_i w_j = \sum_{j \neq i} a_{ij} w_i w_j + \sum_{j \neq i} a_{ji} w_j w_i + a_{ii} w_i^2 + \ldots$$

$$\frac{\partial \ldots}{\partial w_i} = \sum_j a_{ij} w_j + \sum_j a_{ji} w_j$$

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

$\cancel{X^TX\bar w = X^T\bar y}$

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



$$p(y_n|\bar w, \bar x_n) = ?$$

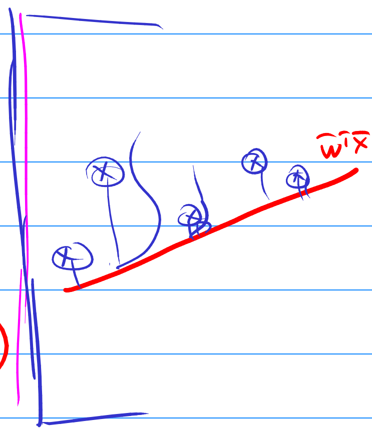
$$y_n = \bar w^T \bar x_n + \varepsilon,$$

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

$D = \{y_n | x_n\}$

$$p(\bar w|D) \propto p(\bar w) \cdot p(D|\bar w) = $$

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



$\bar w_{ML} = ?$ $\qquad p(D|\bar w) \xrightarrow[\bar w]{} \max$

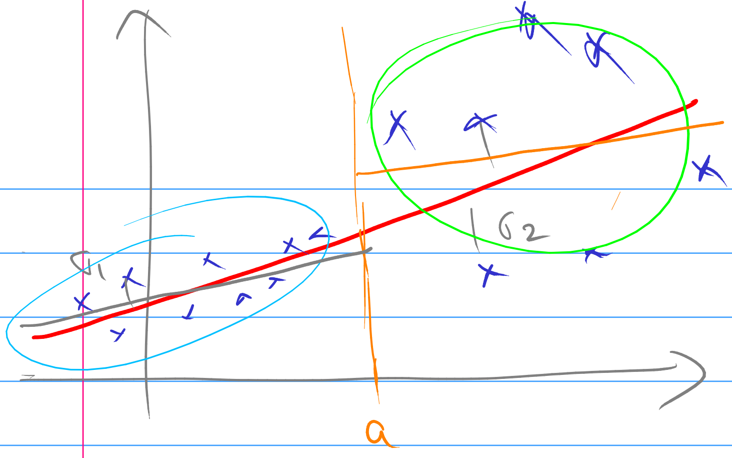
$$p(D|\bar w) = \prod_{n=1}^{N} p(y_n|\bar w, \bar x_n) = \prod_{n=1}^{N} \mathcal{N}(y_n|\bar w^T \bar x_n, \sigma^2) =$$

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

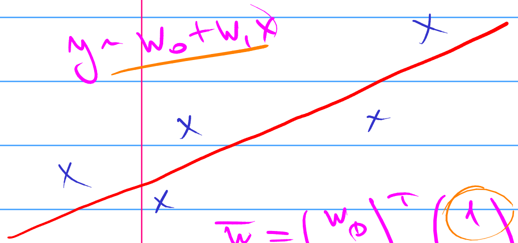
$$\ln p(D|\bar w) = \cancel{-\frac{N}{2}\ln(2\pi\sigma^2)} - \sum_{n=1}^{N} \frac{1}{2\sigma^2}(y_n - \bar w^T \bar x_n)^2 \xrightarrow[\bar w]{} \max$$

$$\Updownarrow$$

$$\sum_{n=1}^{N} (y_n - \bar w^T \bar x_n)^2 \xrightarrow[\bar w]{} \min$$

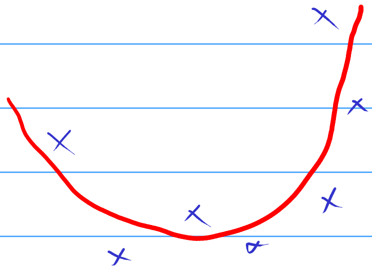


$$\sum_{x_n<a} (y_n - \bar{x}_n^T \bar{w})^2 + \sum_{x_n>a} \frac{\sigma_1^2}{\sigma_2^2} (y_n - \bar{x}_n^T \bar{w})^2 \to \min_{\bar{w}}$$



$y \sim w_0 + w_1 x$

$$\bar{w} = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}^T \begin{pmatrix} 1 \\ x \end{pmatrix} \sim y$$



$y \sim w_0 + w_1 x + w_2 x^2$

$$\begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix}^T \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix} \sim y$$

$$y \sim \bar{w}^T \bar{x}$$

$$y \sim \bar{w}^T \bar{x} + w_0 \iff y \sim \begin{pmatrix} \bar{w} \\ w_0 \end{pmatrix}^T \begin{pmatrix} \bar{x} \\ 1 \end{pmatrix}$$

$$p(y_n \mid \bar{w}, \bar{x}_n) = \mathcal{N}\left(y_n \,\middle|\, \bar{w}^T \begin{pmatrix} 1 \\ x_n \\ x_n^2 \end{pmatrix}, \sigma^2\right)$$

$$\bar{x} \longrightarrow \boxed{\bar{\varphi}} \longrightarrow \bar{\varphi}(\bar{x}) \longrightarrow \boxed{\text{Lin.Regr.}}$$

RBF – radial basis functions