напри. data prior likelihood

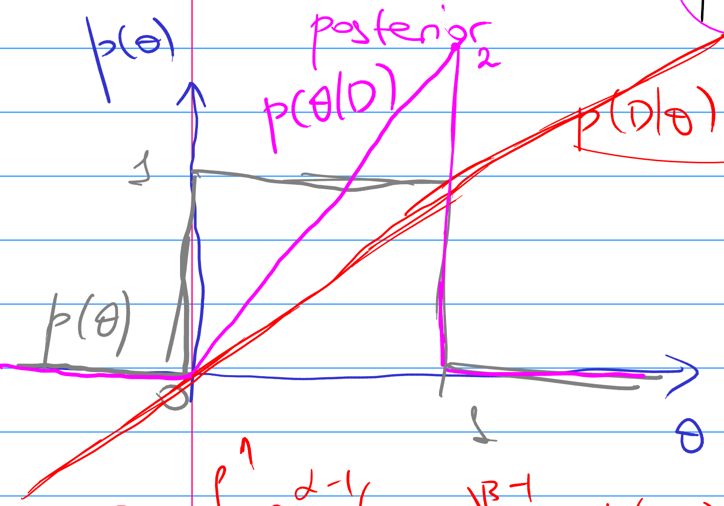
$$p(\theta|D) = \frac{p(\theta)\,p(D|\theta)}{p(D)}$$

posterior

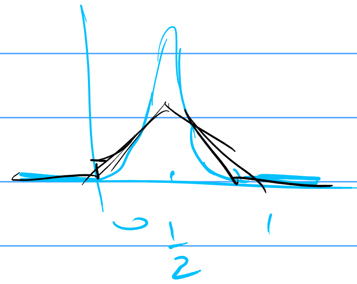
$$p(D|\theta) \xrightarrow[\theta]{} \max$$

$$p(\theta|D), \quad p(\theta|D) \xrightarrow[\theta]{} \max$$

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



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



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

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

$$\begin{cases} \frac{1}{Z}\theta^{\alpha-1}(1-\theta)^{\beta-1}, & \theta\in[0,1] \\ 0, & \theta\notin[0,1] \end{cases} \times \theta^{n}(1-\theta)^{m} \propto \begin{cases} \theta^{n+\alpha-1}(1-\theta)^{m+\beta-1}, & \theta\in[0,1] \\ 0, & \theta\notin[0,1] \end{cases}$$

Beta distr.

Beta(α, β)

$\times \quad D=(n,m)$

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

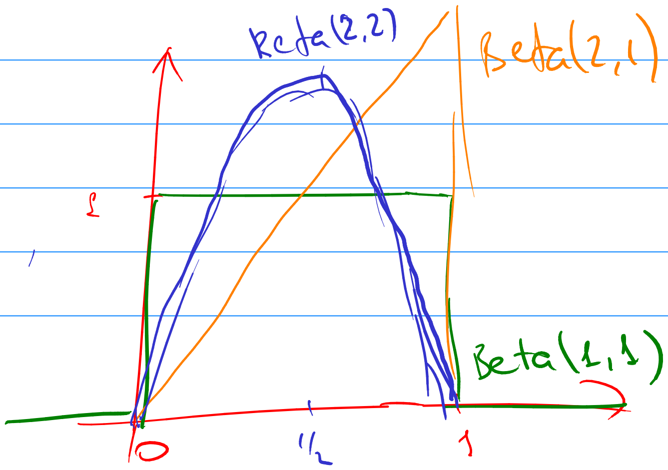
conjugate prior

Beta distr

Beta(n+α, m+β)

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

$$\times \quad \frac{1}{Z}\theta^{n}(1-\theta)^{m} e^{-\,\cdots}$$



Beta(2,1) $\propto \theta$, $[0,1]$

$$\mathrm{Beta}\left(\theta \,\middle|\, \tfrac{1}{2}, \tfrac{1}{2}\right) \propto$$

$$\propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}$$

$$= \frac{1}{\sqrt{\theta(1-\theta)}}$$

$$\mathrm{Beta}(\alpha,\beta) \times D=(n,m) = \mathrm{Beta}(\alpha+n,\beta+m) \times D'=(n',m') = \mathrm{Beta}(\alpha+n+n',\ \beta+m+m')$$

$$\hat{D} = (n+n',\ m+m')$$

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

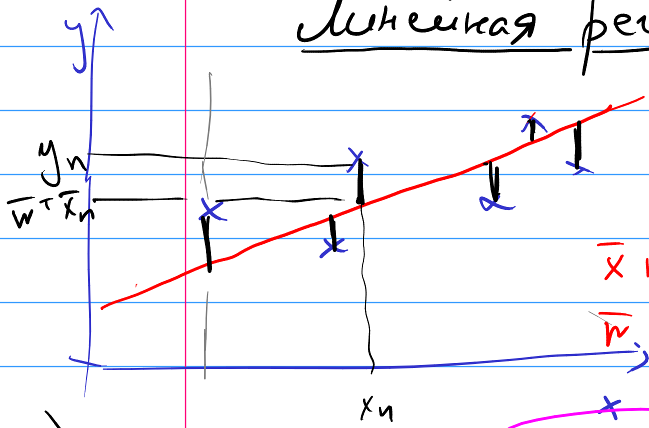
$$p(\theta_1,\dots,\theta_k) = \begin{cases} \frac{1}{Z}\theta_1^{\alpha_1-1}\theta_2^{\alpha_2-1}\dots\theta_k^{\alpha_k-1}, & \sum\theta_i = 1,\ \theta_i \ge 0 \\ 0, & \sum\theta_i \ne 1 \end{cases}$$

$= \mathrm{Dir}(\alpha_1,\dots,\alpha_k)$

Dirichlet distr.

$$p(\bar\theta|\bar\alpha) = \mathrm{Dir}\left(\bar\theta \,\Big|\, \frac{1}{10},\dots,\frac{1}{10}\right) \longrightarrow \text{sparsity}$$

Линейная регрессия



$w_1x + w_0$

$\bar w^T\bar x$

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

$\bar w \mapsto \begin{pmatrix}\bar w\\ w_0\end{pmatrix}$

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

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

$\cancel{p(x)}$

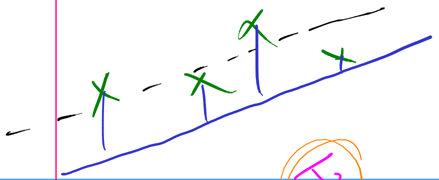
$$y \sim \bar w^T\bar x + \varepsilon \qquad \boxed{\varepsilon \sim \mathcal{N}(\varepsilon|0,\sigma^2)}$$

$$p(y|\bar x,\bar w) = \mathcal{N}(y|\bar w^T\bar x, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(y_n-\bar w^T\bar x_n)^2}$$

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

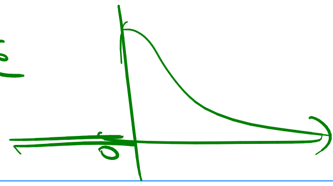
$$\ln p(D|\bar w) = \sum_{n=1}^N\left[-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y_n-\bar w^T\bar x_n)^2\right] =$$

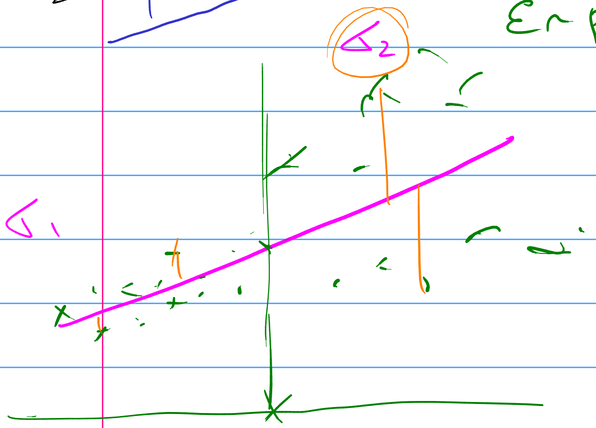
$$= \mathrm{const} - \frac{1}{2\sigma^2}\underbrace{\sum_{n=1}^N (y_n-\bar w^T\bar x_n)^2}_{\xrightarrow[\bar w]{} \min} \xrightarrow[\bar w]{} \max$$



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

$\varepsilon \sim$ Poisson





$$\ln p(D|\bar{w}) = \text{const} - \frac{1}{2}\sum_{n=1}^{N} \frac{1}{\sigma_n^2}\left(y_n - \bar{w}^T\bar{x}_n\right)^2$$

$$y = w_0 + w_1 x + w_2 x^2 + \varepsilon \qquad \varepsilon \sim \mathcal{N}(\varepsilon|0,\sigma^2)$$

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

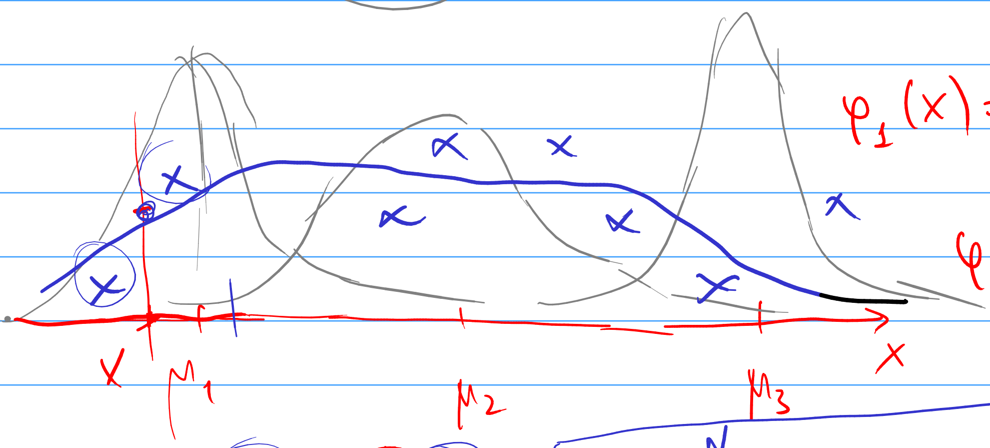
φ

$$\bar{\varphi}(x) = \begin{pmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^d \end{pmatrix}$$

$$\bar{x} \xmapsto{\varphi} \bar{\varphi}(\bar{x})$$

RBF

radial basis functions



$$\varphi_1(x) = e^{-\frac{(x-\mu_1)^2}{\sigma_1^2}}$$

$$\varphi_2(x) = e^{-\frac{(x-\mu_2)^2}{\sigma_2^2}}$$

$$y \approx \begin{pmatrix} \varphi_1(x) \\ \varphi_2(x) \\ \varphi_3(x) \end{pmatrix}^T \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$$

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

$$\begin{pmatrix} \vdots \\ y_n - \bar{w}^T \bar{x}_n \\ \vdots \end{pmatrix} = \underset{N\times 1}{\begin{pmatrix} \vdots \\ y_n \\ \vdots \end{pmatrix}} - \underset{N\times d}{\begin{pmatrix} \\ - \bar{x}_n - \\ \\ \end{pmatrix}} \underset{d\times 1}{\begin{pmatrix} \vdots \\ w_i \\ \vdots \end{pmatrix}} = \bar{y} - X\bar{w}$$

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

$$= \cancel{\bar{y}^T\bar{y}} - \bar{y}^T X \bar{w} - \bar{w}^T X^T \bar{y} + \bar{w}^T X^T X \bar{w} \to \min$$

$$-2\bar{w}^T(X^T\bar{y}) + \bar{w}^T(X^T X)\bar{w} \to \min$$

$$\nabla_{\bar z}\left(\bar a^T \bar z\right) = \bar a \qquad \nabla_{\bar z}\left(\underbrace{\bar z^T \bar z}\right) = 2\bar z$$

$$\nabla_{\bar z} f = \left(\ldots \frac{\partial f}{\partial z_i} \ldots\right) \qquad \frac{\partial\left(\sum_i z_i^2\right)}{\partial z_j} = 2 z_j$$

$$\nabla_{\bar z}\left(\underbrace{\bar z^T A \bar z}\right) = \left(\ldots, \sum_j a_{kj} z_j + \sum_i a_{ik} z_i, \ldots\right) = A\bar z + A^T \bar z = (A + A^T)\bar z$$

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

$$= \sum_{j\neq k} a_{kj} z_j + \sum_{i\neq k} a_{ik} z_i + 2 a_{kk} z_k =$$

$$= \underbrace{\sum_j a_{kj} z_j} + \sum_i a_{ik} z_i$$

$$A = \left(\; a_{ij} \;\right)$$

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

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

$$\boxed{\bar w_{ML} = (X^T X)^{-1} X^T \bar y}$$

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

$$p(\bar w) = \mathcal N(\bar w | \bar 0, \sigma_0^2 I)$$

max

$$\uparrow_{\bar w}\; p(\bar w | D) \propto p(\bar w)\, p(D|\bar w)$$

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

$$- \cancel{\frac{d}{2}\ln(2\pi\sigma_0^2)} - \frac{1}{2\sigma_0^2}\bar w^T \bar w =$$

$$= \text{const} - \frac{1}{2\sigma^2}\left(\underbrace{\frac{1}{\cancel{\sigma^2}}\sum_{n=1}^{N}\left(y_n - \bar w^T \bar x_n\right)^2 + \frac{\sigma^2}{\sigma_0^2}\bar w^T \bar w}\right)$$

$$\Rightarrow \min_{\bar w}$$

$$\boxed{\mathcal N(\bar x | \bar\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} \cdot \sqrt{\det \Sigma}} \cdot e^{-\frac{1}{2}(\bar x - \bar\mu)^T \Sigma^{-1} (\bar x - \bar\mu)}}$$

ridge regression

$$\sum_{n=1}^{N}\left(y_n-\bar{w}^T\bar{x}_n\right)^2+\underbrace{\left(\frac{\sigma^2}{\sigma_0^2}\right)^{=\lambda}\cdot\bar{w}^T\bar{w}}_{\text{regularizer}}\xrightarrow[\bar{w}]{}\min$$

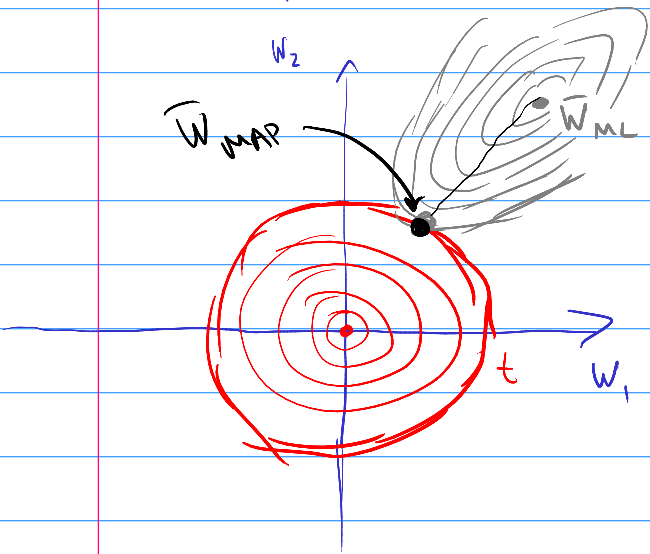
lasso regression

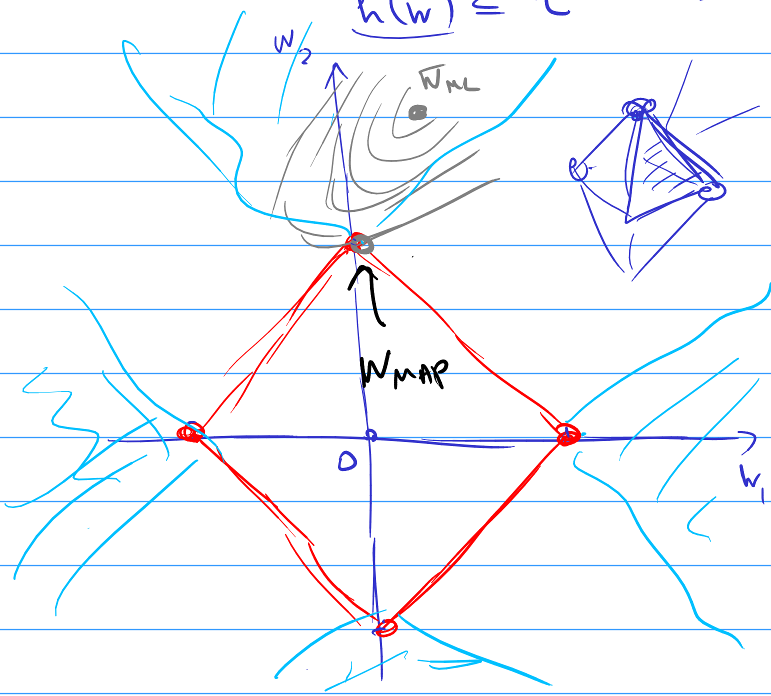
$$\sum_{n=1}^{N}\left(y_n-\bar{w}^T\bar{x}_n\right)^2+\lambda\cdot\underbrace{\sum_{i=1}^{d}|w_i|}\xrightarrow[\bar{w}]{}\min$$

$$p(\bar{w})=\dots e^{-\dots\sum|w_i|}$$

$$F(\bar{w})+\lambda h(\bar{w})\to\min \iff \begin{matrix}F(\bar{w})\to\min\\ h(\bar{w})\le t\end{matrix}$$

$\sum|w_i|=t$





classif. $C_1 \sim C_K$

regression $y \in \mathbb{R}$

Learning to rank