posterior $p(\theta;\alpha')$ — conjugate prior

$$p(\theta|D)=\frac{p(\theta)\,p(D|\theta)}{p(D)}, \qquad p(D|\theta) = \prod N(y_n|\bar w^T\bar x_n,\sigma^2)$$

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

$y\approx \bar x^T\bar w+\varepsilon_{\sim N(0,\sigma^2)}$

$$W_{MAP}=\arg\min\Big[\underbrace{\sum_n(\bar w^T\bar x_n-y_n)^2}+\lambda\cdot\underbrace{\sum w_i^2}_{=\bar w^T\bar w}\Big]$$

$y\approx\bar\varphi(\bar x)^T\bar w+\varepsilon$

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

$p(\bar w)=N(\bar w|\bar 0,\lambda I)$

$$\ln p(\bar w|D)=\text{const}-\frac{1}{2}\cdot\bar w^T\bar w-\underbrace{\frac{1}{2\sigma^2}\sum_n(\bar w^T\bar x_n-y_n)^2}\longrightarrow\max$$

$$p(\theta)=B(\alpha,\beta)\times\underbrace{\theta^{n_h}(1-\theta)^{n_t}}\propto B(\alpha+n_h,\beta+n_t)$$

$$\boxed{p(\bar w|\mu_0,\Sigma_0)}=\frac{1}{(2\pi)^{d/2}\cdot\sqrt{\det\Sigma_0}}e^{-\frac{1}{2}(\bar w-\bar\mu_0)^T\Sigma_0^{-1}(\bar w-\bar\mu_0)}$$

$\times$

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

$$\ln p(\bar w|D)=\ln p(\bar w|\mu_0,\Sigma_0)+\ln p(D|\bar w)+\text{const}=$$

$$=-\cancel{\frac{d}{2}\ln 2\pi}-\cancel{\frac{1}{2}\ln\det\Sigma_0}-\frac{1}{2}(\bar w-\bar\mu_0)^T\Sigma_0^{-1}(\bar w-\bar\mu_0)-\cancel{\frac{N}{2}\ln 2\pi\sigma^2}-\frac{1}{2\sigma^2}\sum_{n=1}^N(y_n-\bar x_n^T\bar w)^2+\text{const}$$

$$\sum_{n=1}^N(y_n-\bar x_n^T\bar w)^2:\quad -\frac{1}{2\sigma^2}(\bar y-X\bar w)^T(\bar y-X\bar w) = -\frac{1}{2\sigma^2}(\bar y^T\bar y-2\bar w^TX^T\bar y+\bar w^TX^TX\bar w)$$

$$\ln p(\bar w|\mu_N,\Sigma_N)=\text{const}-\frac{1}{2}(\bar w-\bar\mu_N)^T\Sigma_N^{-1}(\bar w-\bar\mu_N)$$

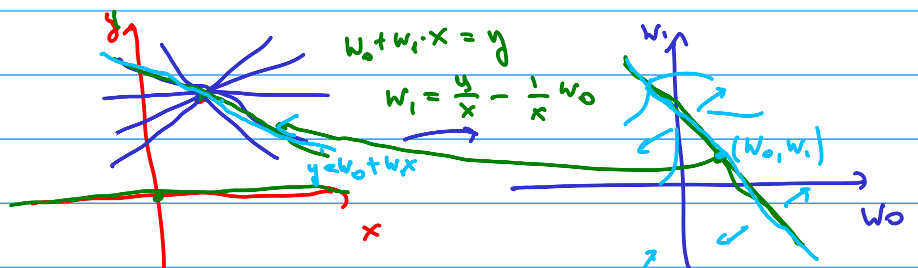
$$-\frac{1}{2}\bar\mu_N^T\Sigma_N^{-1}\mu_N-\frac{1}{2}\bar w^T\Sigma_N^{-1}\bar w+\bar w^T\boxed{\Sigma_N^{-1}\bar\mu_N}$$

$$-\frac{1}{2}\bar w^T\Sigma_0^{-1}\bar w^T-\frac{1}{2\sigma^2}\bar w^TX^TX\bar w=-\frac{1}{2}\bar w^T\underbrace{\Big[\Sigma_0^{-1}+\frac{1}{\sigma^2}X^TX\Big]}_{=\Sigma_N^{-1}}\bar w$$

$$\Sigma_N=\Big[\Sigma_0^{-1}+\frac{1}{\sigma^2}X^TX\Big]^{-1}$$

$$\bar w^T\cdot\Sigma_0^{-1}\bar\mu_0+\frac{1}{\sigma^2}\bar w^TX^T\bar y=\bar w^T\underbrace{\Big(\Sigma_0^{-1}\bar\mu_0+\frac{1}{\sigma^2}X^T\bar y\Big)}_{=\Sigma_N^{-1}\mu_N}$$

$$\mu_N=\Sigma_N\cdot\Big[\Sigma_0^{-1}\bar\mu_0+\frac{1}{\sigma^2}X^T\bar y\Big]$$



log likelihood — log prior — ridge regression

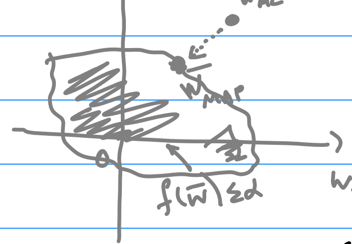
$$\sum\left(\bar w^T \bar x_n - y_n\right)^2 + \frac{\lambda}{2}\,\bar w^T \bar w \longrightarrow \min$$

$$p(\bar w) \propto e^{-\alpha\cdot\Sigma|w_i|}$$

Lasso regression

$$\sum_{n=1}^{N}\left(\bar w^T \bar x_n - y_n\right)^2 + \alpha\cdot\sum_{i=1}^{d}|w_i| \longrightarrow \min$$

$$\bar w_{ML} = \arg\min \Sigma\left(\bar w^T \bar x_n + y_n\right)^2$$



$$\sum^{N}\left(\bar w^T \bar x_n - y_n\right)^2 + \frac{\lambda}{2}\cdot f(\bar w) \longrightarrow \min$$

$$\Updownarrow$$

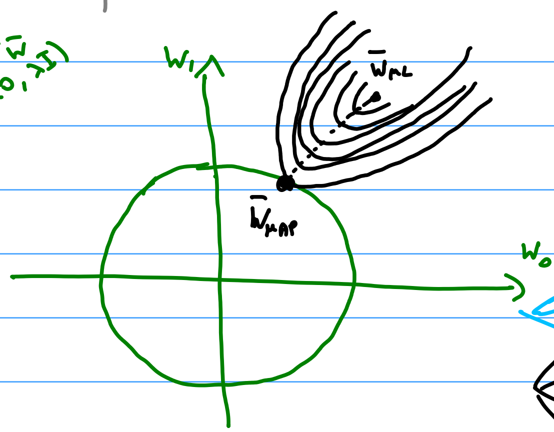
$$\Sigma\left(\bar w^T \bar x_n - y_n\right)^2 \longrightarrow \min$$

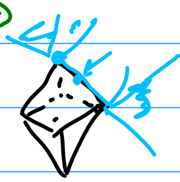
$$f(\bar w) \le \alpha$$

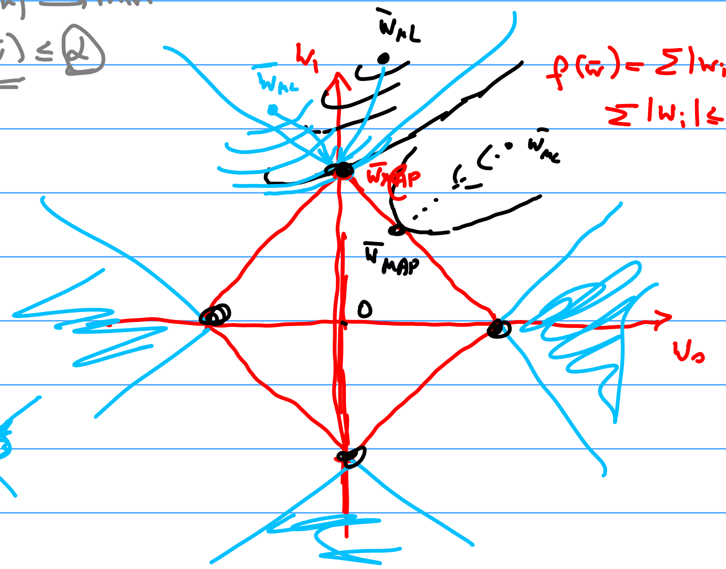
$f(\bar w) = \bar w^T \bar w$

$p(\bar w) = \mathcal{N}(0, \lambda I)$

$\|\bar w\|^2 \le \alpha$







$p(\bar w) = \Sigma|w_i|$

$\Sigma|w_i| \le \alpha$

$\bar x = (x_1 \dots x_{100})$

$\bar w = (w_1 \dots w_{100})$

$\boxed{w_i,\ w_j}$

$\boxed{y \approx w_i x_i + w_j x_j}$

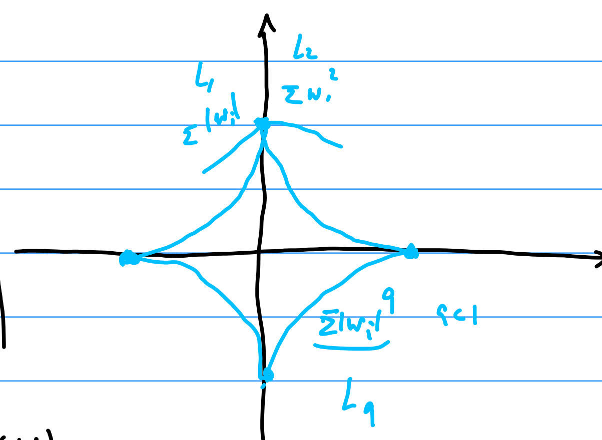
$\boxed{\Sigma\left(\bar x_n^T \bar w - y\right)^2 \rightarrow \min}$

for $i,j = 1 \dots d$

[. —]

$$i^*, j^* = \underset{i,j,w_i,w_j}{\arg\min}\left[\Sigma\left(x_i w_i + x_j w_j - y_n\right)^2\right]$$

$O(d^2)$



$w_1 - w_{100}$ $\quad (w_0, w_{i^*})$

$w_1 \dots w_{100}$ $\quad (w_0, w_{i^*}, w_{j^*}) \dots$ $\quad O(d\cdot k)$

$$p(\bar w|D) \propto p(\bar w)\,p(D|\bar w)$$

$$p(y|\bar x, D) = \int p(y|\bar x, \bar w)\cdot p(\bar w|D)\,d\bar w$$

$$\propto \int p(y|\bar x,\bar w)\,p(\bar w)\,p(D|\bar w)\,d\bar w$$

$$\int \mathcal{N}\left(y \mid \bar w^T \bar x, \sigma^2\right)\cdot\mathcal{N}\left(\bar w \mid \mu_N, \Sigma_N\right)d\bar w$$

$$\ln_{\dots} = -\frac{1}{2}\cancel{\ln(2\pi\sigma^2)} - \frac{1}{2\sigma^2}\left(y - \bar w^T \bar x\right)^2 - \frac{d}{2}\cancel{\ln 2\pi} - \cancel{\frac{1}{2}\ln\det\Sigma_N} - \frac{1}{2}(\bar w - \mu_N)^T\Sigma_N^{-1}(\bar w - \mu_N)$$

$$-\frac{1}{2}\cdot\frac{1}{\sigma^2}\cdot \bar w^T \underbrace{\bar x\cdot x^T}_{d\times 1\cdot 1\times d}\cdot\bar w - \frac{1}{2}\bar w^T\Sigma_N^{-1}\cdot\bar w$$

$$\int c(y)\cdot\mathcal{N}\left(\bar w \mid \bar\mu', \Sigma'\right)d\bar w = c(y)$$

$$\Sigma'^{-1} = \Sigma_N^{-1} + \frac{1}{\sigma^2}\bar x\,\bar x^T$$

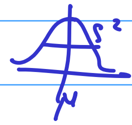
$\boxed{Д/З}$: 1) μ'

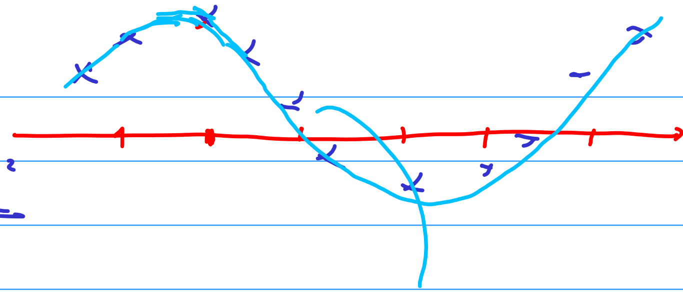
2) $c(y) = \dots y \dots y^2$

3) $p(y|\bar x, D) = \mathcal{N}(y \mid \mu', \sigma'^2)$

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

$$\varphi(x; s, \mu) = e^{-\frac{1}{2s^2}(x-\mu)^2}$$





$$\varphi_j(x; s, \mu_j)$$