Empirical Bayes

$1/\sigma^2 = \beta$ $\qquad = \sqrt{\frac{\beta}{2\pi}}\, e^{-\frac{\beta}{2}(y-\bar{x}^T\bar{w})^2}$

Model selection

$$p(y|\bar{x},\bar{w}) = \underline{N(y|\bar{x}^T\bar{w}, \beta)}$$

$$\frac{\alpha}{\beta}$$

$$p(\bar{w}|\bar{0},\alpha I) = \underbrace{\prod_i N(w_i|0,\alpha)}$$

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

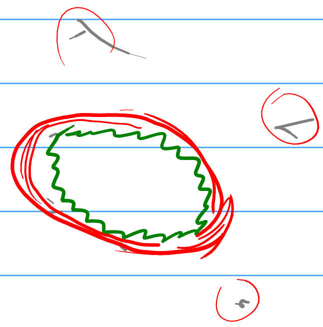
$$p(D|\alpha,\beta) \xrightarrow[\alpha,\beta]{} \max$$

$$\parallel$$

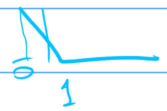
$$\int p(\bar{w}|\alpha)\, p(D|\bar{w},\beta)\, d\bar{w}$$

$$\parallel$$

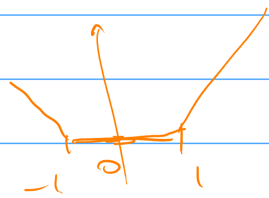
$$\int \underbrace{\left(\frac{\alpha}{2\pi}\right)^{D/2} \left(\frac{\beta}{2\pi}\right)^{N/2}} e^{-\overbrace{\left(\frac{\alpha}{2}\bar{w}^T\bar{w} + \frac{\beta}{2}\|\bar{y}-X\bar{w}\|^2\right)}^{=E(\bar{w})}}\, d\bar{w}$$



$$E(\bar{w}) = \underbrace{\frac{1}{2}(\bar{w}-\bar{m})^T A (\bar{w}-\bar{m})}_{A\,=\,\alpha I + \beta X^T X} + \qquad \bar{m} = \beta A^{-1} X^T \bar{y}$$



$$\underbrace{+ \frac{\beta}{2}\|\bar{y} - X\bar{m}\|^2 + \frac{\alpha}{2}\bar{m}^T\bar{m}}$$



$$p(D|\alpha,\beta) = \left(\frac{\alpha}{\cancel{2\pi}}\right)^{D/2}\left(\frac{\beta}{2\pi}\right)^{N/2} \frac{(\cancel{2\pi})^{\cancel{D/2}}}{\sqrt{\det A}} \cdot e^{-}$$

$$\ln p(D|\alpha,\beta) = \frac{D}{2}\ln\alpha + \frac{N}{2}\ln\beta - \frac{1}{2}\underline{\ln\det A} - \underline{\frac{\beta}{2}\|\bar{y}-X\bar{m}\|^2 - \frac{\alpha}{2}\bar{m}^T\bar{m}} - \frac{N}{2}\ln 2\pi$$

$\downarrow \alpha,\beta$

max

Relevance vector machines

① $p(y|\bar{x},\bar{w},\beta) = \mathcal{N}(y|\bar{w}^T\bar{x},\beta)$ $\quad {}^{''}1/\sigma^2$

$$\cancel{p(\bar{w}|\alpha) = \mathcal{N}(\bar{w}|\bar{0},\alpha I)}$$

$$p(\bar{w}|\bar{\alpha}) = \mathcal{N}\left(\bar{w}\,\middle|\,\bar{0}, \underbrace{\begin{pmatrix}\alpha_1 & & \\ & \ddots & \\ & & \alpha_D\end{pmatrix}}_{=A}\right) = \prod_{i=1}^{D} \mathcal{N}(w_i|0,\alpha_i)$$

$$\underbrace{\ln p(D|\bar{\alpha},\beta)}_{} = \frac{1}{2}\sum \ln\alpha_i + \frac{N}{2}\ln\beta - \frac{1}{2}\ln\det\Sigma - \frac{N}{2}\ln 2\pi - \frac{\beta}{2}\|\bar{y}-X\bar{m}\|^2 - \frac{1}{2}\bar{m}^T\begin{pmatrix}\alpha_1 & & \\ & \ddots & \\ & & \alpha_D\end{pmatrix}\bar{m}$$

$$\xrightarrow{\alpha_i,\,\beta} \max$$

$$\Sigma = (A + \beta X^T X)^{-1}$$

$$\bar{m} = \beta \Sigma X^T \bar{y}$$

$\bar{x} = f(\bar{x})$

$$\boxed{\alpha_i = \frac{\gamma_i}{m_i^2}, \quad \beta = \frac{N - \sum_i \gamma_i}{\|\bar{y} - X\bar{m}\|^2}}$$

$$\gamma_i = 1 - \alpha_i \Sigma_{ii} \longrightarrow \boxed{\alpha_i^*, \beta^*}$$

∞

$$p(y|\bar{x}, D) = \mathcal{N}(y|\bar{m}^T\bar{x}, \beta^{-1} + \bar{x}^T\Sigma\bar{x})$$

$$\boxed{w_i = \mathcal{N}(w_i|0, \alpha_i^*)} \quad \sigma_i^2 \to 0$$

$$\alpha_i^* \to \infty$$