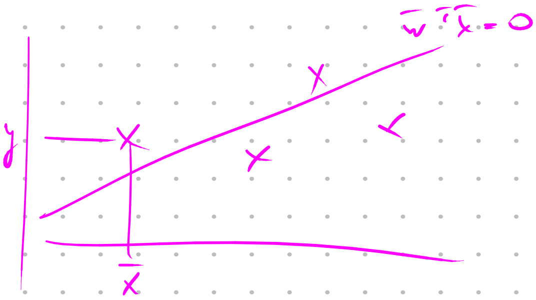


$$p(y \mid \bar{x}, \bar{w}) = \mathcal{N}(y \mid \bar{x}^T \bar{w}, \sigma^2)$$

parameters

precision $\tau = \frac{1}{\sigma^2}$

$$p(\bar{w}) = \mathcal{N}(\bar{w} \mid \bar{0}, \sigma_0^2 \cdot I)$$

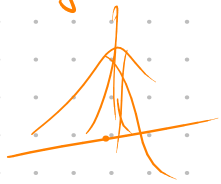
$$\alpha = 1/\sigma_0^2 \qquad \beta = 1/\sigma^2$$

$$\log p(\bar{w} \mid \bar{y}, X) = \text{Const} - \frac{1}{2\sigma^2} \sum_{n=1}^{N} (y_n - \bar{x}_n^T \bar{w})^2 - \frac{1}{2\sigma_0^2} \bar{w}^T \bar{w}$$

$$\log p(\bar{w} \mid D, \alpha, \beta) = \text{const} - \frac{\beta}{2} \sum_n (y_n - \bar{x}_n^T \bar{w})^2 - \frac{\alpha}{2} \bar{w}^T \bar{w}$$

hyperparameters

$$-2 \log p(D \mid \bar{w}_{ML}, \alpha, \beta) + M \log N$$



Empirical Bayes

MLE - II

$$p(\bar{w} \mid D, \alpha, \beta) = \frac{p(\bar{w} \mid \alpha)\, p(D \mid \bar{w}, \beta)}{p(D \mid \alpha, \beta)} \xrightarrow[\alpha, \beta]{} \max$$

evidence function
marginal likelihood

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

Fully Bayesian ...

$$p(\alpha, \beta) \cdot p(D \mid \alpha, \beta) \propto p(\alpha, \beta \mid D)$$

$$p(\bar{y} \mid X, \alpha, \beta) = \int p(\bar{w} \mid \alpha)\, p(\bar{y} \mid X, \bar{w}, \beta)\, d\bar{w} =$$

$$= \int \left(\frac{\alpha}{2\pi}\right)^{d/2} e^{-\frac{\alpha}{2} \bar{w}^T \bar{w}} \cdot \left(\frac{\beta}{2\pi}\right)^{N/2} \cdot e^{-\frac{\beta}{2} \sum_n (y_n - \bar{w}^T x_n)^2}\, d\bar{w}$$

$$E(\bar{w}) = -\frac{\alpha}{2} \bar{w}^T \bar{w} - \frac{\beta}{2} (\bar{y} - X\bar{w})^T (\bar{y} - X\bar{w}) =$$

$$= -\frac{\alpha}{2} \bar{w}^T \bar{w} - \frac{\beta}{2} (\bar{y}^T \bar{y} - 2 \bar{w}^T X^T \bar{y} + \bar{w}^T X^T X \bar{w}) =$$

$$= -\frac{1}{2} \bar{w}^T (\alpha \cdot I + \beta \cdot X^T X) \bar{w} + \underbrace{\beta \cdot \bar{w}^T X^T \bar{y}}_{= A \bar{\mu}_N} - \frac{\beta}{2} \bar{y}^T \bar{y} =$$

$$= \boxed{-\frac{1}{2} (\bar{w} - \bar{\mu}_N)^T \cdot A (\bar{w} - \bar{\mu}_N)} - \frac{\beta}{2} \bar{y}^T \bar{y} + \frac{1}{2} \bar{\mu}_N^T A \bar{\mu}_N, \quad \text{use}$$

$$A = \alpha I + \beta \cdot X^T X = \Sigma_N^{-1}, \quad \bar{\mu}_N = \beta \cdot A^{-1} X^T \bar{y}$$

$X^T y = \frac{1}{\beta} A \bar{\mu}_N$

$$\int e^{-\frac{1}{2}(\_\_)^T A(\_\_)} d\bar{w} = \sqrt{\frac{(2\pi)^d}{\det A}}$$

$$p(y \mid X, \alpha, \beta) = \left(\frac{\alpha}{\cancel{2\pi}}\right)^{d/2} \left(\frac{\beta}{2\pi}\right)^{N/2} \cdot \sqrt{\frac{\cancel{(2\pi)^d}}{\det A}} \cdot e^{\frac{1}{2}\bar{\mu}_N^T A \bar{\mu}_N - \frac{\beta}{2}\bar{y}^T\bar{y}} \xrightarrow[\alpha,\beta]{} \max$$

$$\log p(y \mid X, \alpha, \beta) = \frac{d}{2}\log\alpha + \frac{N}{2}\log\beta - \frac{N}{2}\log 2\pi - \frac{1}{2}\log\det A \underbrace{- \frac{\beta}{2}\bar{y}^T\bar{y} + \frac{1}{2}\bar{\mu}_N^T A \bar{\mu}_N}_{} \xrightarrow[\alpha,\beta]{} \max$$

$$E(\bar{w}) = -\frac{\beta}{2}(\bar{y} - X\bar{w})^T(\bar{y} - X\bar{w}) - \frac{\alpha}{2}\bar{w}^T\bar{w}$$

$$\underbrace{-\frac{\beta}{2}\bar{y}^T\bar{y}} + \frac{1}{2}\alpha\,\bar{\mu}_N^T\bar{\mu}_N + \underbrace{\frac{\beta}{2}\bar{\mu}_N^T X^T X \bar{\mu}_N} =$$

$$= -\frac{\beta}{2}\left(\bar{\mu}_N^T X^T X \bar{\mu}_N - \underbrace{2\bar{\mu}_N^T X^T \bar{y}} + \bar{y}^T\bar{y}\right) + \frac{\alpha}{2}\bar{\mu}_N^T\bar{\mu}_N + \beta\,\bar{\mu}_N^T X^T X \bar{\mu}_N \underbrace{- \beta\,\bar{\mu}_N^T X^T \bar{y}}_{= \bar{\mu}_N^T A \bar{\mu}_N}$$

$$= -\frac{\beta}{2}(\bar{y} - X\bar{\mu}_N)^T(\bar{y} - X\bar{\mu}_N) - \cancel{\bar{\mu}_N^T A \bar{\mu}_N} +$$

$$+ (\alpha - \frac{\alpha}{2})\bar{\mu}_N^T\bar{\mu}_N + \cancel{\beta\,\bar{\mu}_N X^T X \bar{\mu}_N} =$$

$$= -\frac{\beta}{2}(\bar{y} - X\bar{\mu}_N)^T(\bar{y} - X\bar{\mu}_N) - \frac{\alpha}{2}\bar{\mu}_N^T\bar{\mu}_N = E(\bar{\mu}_N)$$

$$E(\bar{w}) = E(\bar{\mu}_N) - \frac{1}{2}(\bar{w} - \bar{\mu}_N)^T A (\bar{w} - \bar{\mu}_N)$$

$$\log p(D \mid \alpha, \beta) = E(\bar{\mu}_N) + \frac{d}{2}\log\alpha + \frac{N}{2}\log\beta - \frac{N}{2}\log 2\pi - \frac{1}{2}\log\det A$$

$$-\frac{\beta}{2}\bar{y}^T\bar{y} + \frac{1}{2}\bar{\mu}_N^T A \bar{\mu}_N =$$

$$= -\beta/2\,\bar{y}^T\bar{y} + \frac{\alpha}{2}\bar{\mu}_N^T\bar{\mu}_N + \frac{\beta}{2}\bar{\mu}_N^T X^T X \bar{\mu}_N$$

$$A = \alpha I + \beta \cdot (X^T X)$$

λ_i — собств. числа $\beta(X^T X)$: $\quad \beta X^T X \cdot \bar{u}_i = \lambda_i \bar{u}_i$

$$(\alpha I + \beta X^T X)\bar{u}_i = (\lambda_i + \alpha)\bar{u}_i$$

$\lambda_i + \alpha$ — с.ч. матр. A

$$\log\det A = \sum_{i=1}^{d} \log(\lambda_i+\alpha)$$

(*) почему?

$$\frac{\partial \log p(D|\alpha,\beta)}{\partial\alpha} = \frac{d}{2\alpha} - \frac{1}{2}\bar\mu_N^T\bar\mu_N - \frac{1}{2}\sum_i \frac{1}{\lambda_i+\alpha} = 0$$

$$\alpha\cdot\bar\mu_N^T\bar\mu_N + \alpha\cdot\sum_i\frac{1}{\lambda_i+\alpha} = d$$

$$\alpha\cdot(\bar\mu_N^T\bar\mu_N) = d - \alpha\cdot\sum_{i=1}^{d} \frac{1}{\lambda_i+\alpha} = \sum_{i=1}^{d}\left(1-\frac{\alpha}{\lambda_i+\alpha}\right) = \sum_{i=1}^{d} \frac{\lambda_i}{\lambda_i+\alpha}$$

$\alpha^{(0)}$,

$$\alpha^{(k+1)} = \frac{1}{\bar\mu_N(\alpha^{(k)})^T\,\bar\mu_N(\alpha^{(k)})}\cdot\sum_{i=1}^{d}\frac{\lambda_i}{\lambda_i+\alpha^{(k)}}$$

λ_i – с.ч. $\beta(X^TX) \Rightarrow \lambda_i \propto \beta$, $\quad \frac{\partial\lambda_i}{\partial\beta} = D_i = \frac{\lambda_i}{\beta}$

D_i – с.ч. $X^TX \Rightarrow \lambda_i^{(k)} = \beta^{(k)}\cdot D_i$

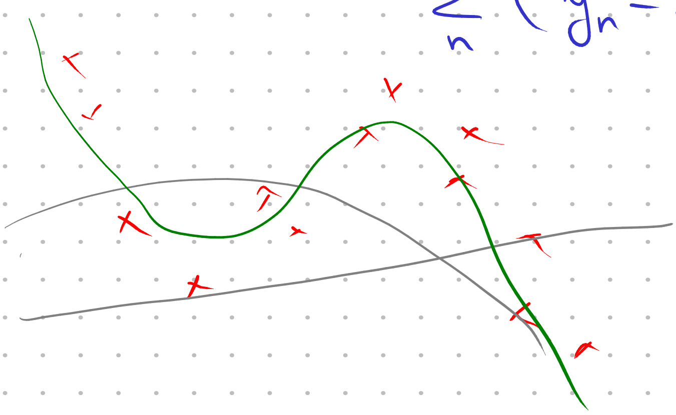
$$\frac{\partial\log\det A}{\partial\beta} = \frac{\partial\left(\sum_i\log(\lambda_i+\alpha)\right)}{\partial\beta} = \sum_i\frac{1}{\lambda_i+\alpha}\cdot\frac{\lambda_i}{\beta} = \frac{1}{\beta}\cdot\sum_i\frac{\lambda_i}{\lambda_i+\alpha}$$

$$\frac{\partial\log p(D|\alpha,\beta)}{\partial\beta} = \frac{N}{\cancel{2}\beta} - \frac{1}{\cancel{2}\beta}\cdot\sum_i\frac{\lambda_i}{\lambda_i+\alpha} - \frac{1}{\cancel{2}}\cdot\underbrace{\sum_n \left(y_n - \bar\mu_N^T x_N\right)^2} = 0$$

$\beta^{(0)}$

$$\beta^{(k+1)} = \frac{N - \sum_i \frac{\lambda_i}{\lambda_i+\alpha^{(k)}}}{\sum_n \left(y_n - \bar\mu_N(\alpha^{(k)},\beta^{(k)})^T \bar x_n\right)^2}$$

$X = (\;)^2$



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

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

$$y \sim w_0 + w_1 x + w_2 x^2 + w_3 x^3 = \begin{pmatrix}w_0\\w_1\\w_2\\w_3\end{pmatrix}^T\begin{pmatrix}1\\x\\x^2\\x^3\end{pmatrix}$$

$X = (\;)^4$