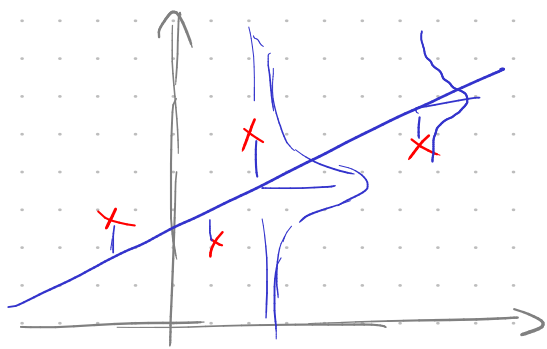




$$y \approx \bar{w}^T \bar{x}$$

$$\sum_n (y_n - \bar{w}^T \bar{x}_n)^2$$

$$x \mapsto \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix}$$

$$x \mapsto \begin{pmatrix} \vdots \\ c_i - c_i(x-\mu)^2 \\ \vdots \end{pmatrix}$$

① Regularization

$$\bar{X} = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$$

$\mathbb{R}^d$

$\mapsto$

$$\begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_d \\ x_1^2 \\ x_2^2 \\ \vdots \\ x_1 x_2 \\ \vdots \\ x_d^2 \end{pmatrix}$$

$$\in \mathbb{R}^{\frac{d(d+1)}{2}}$$

log-likelihood

log-prior

$p(\bar{w})$

$$L(\bar{w}) = \sum_n (\bar{w}^T \bar{x}_n - y_n)^2 + R(\bar{w})$$

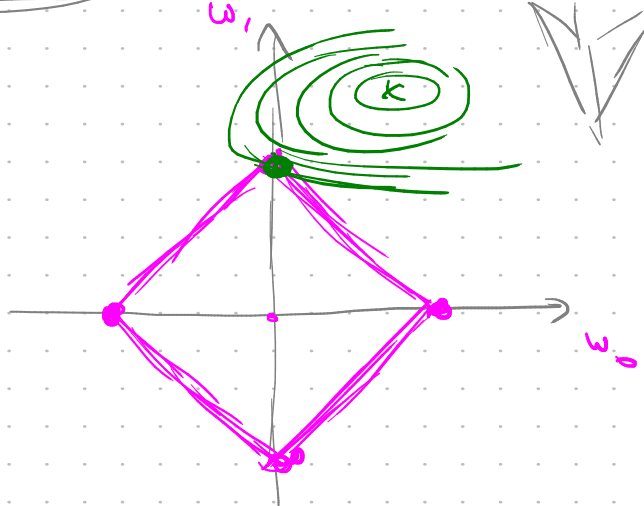
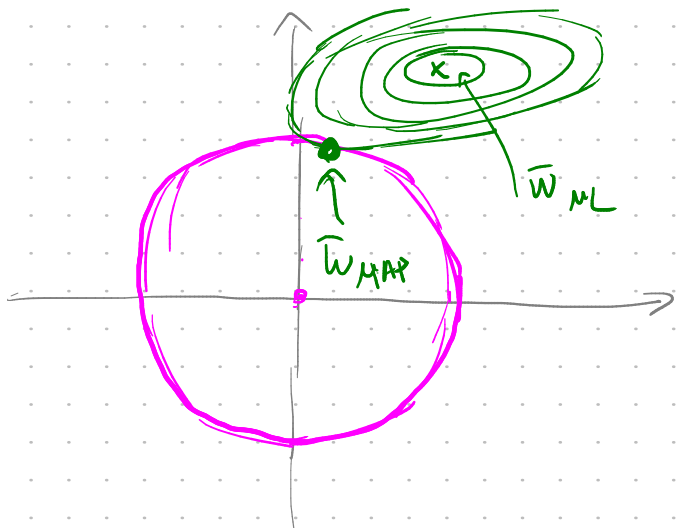
Ridge

LASSO

$$\propto \sum_{i=1}^d w_i^2$$

$$\propto \sum_{i=1}^d |w_i|$$

$$\min \sum_n (-)^2 \text{ s.t. } R(\bar{w}) \leq \dots$$



② Bayesian Inference

$$p(\bar{w}) \times p(D|\bar{w}) \propto p(\bar{w}|D)$$

$$! \times \prod_n \mathcal{N}(y_n | \bar{w}^T \bar{x}_n, \sigma^2) \propto ?$$

$$\prod_i w_i^{\alpha-1} (1-w_i)^{\beta-1} \propto \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_n - \bar{w}^T \bar{x}_n)^2}$$

$$\dots + \left(\text{Const} - \frac{1}{2\sigma^2} \sum_n (y_n - \bar{w}^T \bar{x}_n)^2 \right) = \dots$$

$$p(\bar{w}) = \mathcal{N}(\bar{w} | \bar{0}, \alpha \mathbf{I}) = \prod_i \mathcal{N}(w_i | 0, \sigma_0^2)$$

$$\log p(\bar{w} | D) = \text{const} + \log p(D | \bar{w}) + \log p(\bar{w}) =$$

$$\stackrel{\text{max}}{\downarrow} = \text{const} - \frac{1}{2\sigma^2} \sum_{n=1}^N (y_n - \bar{w}^T \bar{x}_n)^2 - \sum_{i=1}^d \frac{1}{2\sigma_0^2} w_i^2 =$$

$$= \text{const} - \frac{1}{2\sigma^2} \left(\sum_n (y_n - \bar{w}^T \bar{x}_n)^2 + \frac{\sigma^2}{\sigma_0^2} \sum_i w_i^2 \right)$$

$$p(\bar{w}) \propto e^{-c \cdot \sum_i |w_i|} \quad \xrightarrow{\bar{w} \rightarrow \min} \quad \sum_i |w_i|$$

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

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

$$= \dots + \bar{w}^T \underbrace{(X^T X + \alpha \mathbf{I})}_{\text{regularization}} \bar{w}$$

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

$$\boxed{\mathcal{N}(\bar{w} | \bar{\mu}_0, \Sigma_0)} \times \boxed{\prod_n \mathcal{N}(y_n | \bar{w}^T \bar{x}_n, \sigma^2)} \propto$$



$$\bar{w} \in \mathbb{R}^d \quad N(\bar{w} | \bar{\mu}_0, \Sigma_0) = \frac{1}{(2\pi)^{d/2} \cdot \sqrt{\det \Sigma_0}} e^{-\frac{1}{2}(\bar{w} - \bar{\mu}_0)^T \Sigma_0^{-1} (\bar{w} - \bar{\mu}_0)}$$

$$\begin{aligned} \log p(\bar{w} | D) &= \text{const} + \log p(\bar{w}) + \log p(D | \bar{w}) = \\ &= \text{const} - \frac{1}{2}(\bar{w} - \bar{\mu}_0)^T \Sigma_0^{-1} (\bar{w} - \bar{\mu}_0) - \frac{1}{2\sigma^2} (\bar{y} - X\bar{w})^T (\bar{y} - X\bar{w}) = \end{aligned}$$

$$p(\bar{w} | D) = N(\bar{w} | \bar{\mu}_N, \Sigma_N) ; \quad \bar{\mu}_N, \Sigma_N = ?$$

$$\begin{aligned} &= \text{const} - \frac{1}{2} \left(\bar{w}^T \Sigma_0^{-1} \bar{w} - \bar{\mu}_0^T \Sigma_0^{-1} \bar{w} - \bar{w}^T \Sigma_0^{-1} \bar{\mu}_0 + \bar{\mu}_0^T \Sigma_0^{-1} \bar{\mu}_0 \right) \\ &\quad + \frac{1}{\sigma^2} \left(\bar{y}^T \bar{y} - \bar{w}^T X^T \bar{y} - \bar{y}^T X \bar{w} + \bar{w}^T X^T X \bar{w} \right) \\ &= \text{const} - \frac{1}{2} \left(\bar{w}^T \Sigma_0^{-1} \bar{w} - 2 \bar{w}^T (\Sigma_0^{-1} \bar{\mu}_0) + \frac{1}{\sigma^2} \bar{w}^T X^T X \bar{w} - \frac{2}{\sigma^2} \bar{w}^T (X^T \bar{y}) \right) \end{aligned}$$

$$p(\bar{w} | D) = N(\bar{w} | \bar{\mu}_N, \Sigma_N) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Sigma_N}} e^{-\frac{1}{2}(\bar{w} - \bar{\mu}_N)^T \Sigma_N^{-1} (\bar{w} - \bar{\mu}_N)}$$

$$\log p(\bar{w} | D) = \text{const} - \frac{1}{2} (\bar{w} - \bar{\mu}_N)^T \Sigma_N^{-1} (\bar{w} - \bar{\mu}_N) =$$

$$= \text{const} - \frac{1}{2} \left(\bar{w}^T \Sigma_N^{-1} \bar{w} - 2 \bar{w}^T (\Sigma_N^{-1} \bar{\mu}_N) \right)$$

$$\boxed{\Sigma_N^{-1} = \Sigma_0^{-1} + \frac{1}{\sigma^2} X^T X}$$

$$\boxed{\bar{\mu}_N = \Sigma_N \left(\Sigma_0^{-1} \bar{\mu}_0 + \frac{1}{\sigma^2} X^T \bar{y} \right)}$$

$$\Sigma_N^{-1} \bar{\mu}_N = \Sigma_0^{-1} \bar{\mu}_0 + \frac{1}{\sigma^2} X^T \bar{y}$$

$$p(\bar{w}) = \mathcal{N}(\bar{w} | \bar{\mu}_0, \bar{\Sigma}_0)$$

$$\times p(D|\bar{w}) = \prod \mathcal{N}(y_n | \bar{w}^T \bar{x}_n, \sigma^2)$$

$$p(\bar{w}|D) = \mathcal{N}(\bar{w} | \bar{\mu}_N, \bar{\Sigma}_N)$$

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

$$= \int \mathcal{N}(y | \bar{w}^T \bar{x}, \sigma^2) \cdot \mathcal{N}(\bar{w} | \bar{\mu}_N, \bar{\Sigma}_N) d\bar{w} =$$

$$= \underline{\mathbb{E}_{p(\bar{w}|D)} [p(y|\bar{x}, \bar{w})]} \approx \frac{1}{R} \sum_{r=1}^R p(y|\bar{x}, \bar{w}^{(r)})$$

$$\underline{y \text{ e } \bar{w}^{(r)} \sim p(\bar{w}|D)}$$