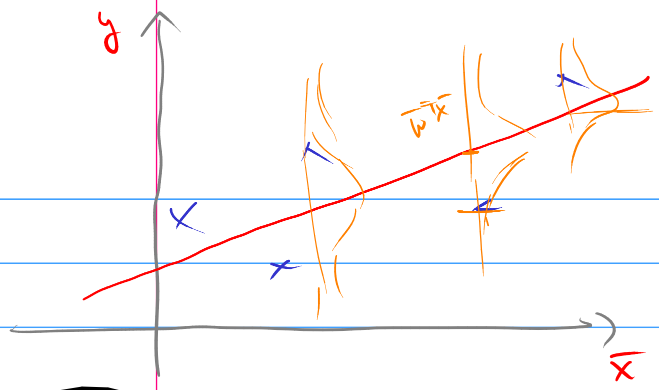


$$y = \bar{w}^T\bar{x} + \varepsilon, \quad \varepsilon \sim N(0,\sigma^2)$$

$$p(D|\bar{w}) = \prod_n p(y_n|\bar{x}_n,\bar{w}) = \prod_n N(y_n|\bar{w}^T\bar{x}_n, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \cdot \prod_{n=1}^{N} e^{-\frac{1}{2\sigma^2}\sum_n (y_n - \bar{w}^T\bar{x}_n)^2} \longrightarrow \max$$

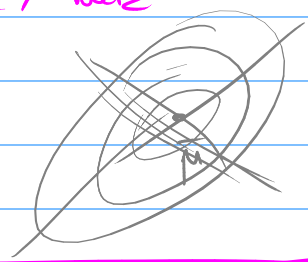
min

Feature selection

$$\underline{p(\bar{w}|D)} = \frac{p(\bar{w})\,p(D|\bar{w})}{p(D)} \propto p(\bar{w}) \cdot \prod_{n=1}^{N} p(y_n|\bar{x}_n,\bar{w})$$

$$\boxed{p(\bar{w}) = N(\bar{w}|\bar{0}, \sigma_0^2 \cdot I)} = \prod_{i=1}^{d} N(w_i|0,\sigma_0^2) = \frac{1}{(2\pi\sigma_0^2)^{d/2}} \cdot e^{-\frac{1}{2\sigma_0^2}\bar{w}^T\bar{w}}$$

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



$$\underline{\ln p(\bar{w}|D)} = \text{Const} - \cancel{\frac{d}{2}\ln(2\pi\sigma_0^2)} - \frac{1}{2\sigma_0^2}\bar{w}^T\bar{w} - \cancel{\frac{N}{2}\ln(2\pi\sigma^2)} - \frac{1}{2\sigma^2}\sum_n (y_n - \bar{w}^T\bar{x}_n)^2$$

$$= \text{Const} - \frac{1}{2\sigma^2}\sum_n (y_n - \bar{w}^T\bar{x}_n)^2 - \frac{1}{2\sigma_0^2}\bar{w}^T\bar{w} \xrightarrow[\bar{w}]{} \max$$

$$\boxed{\sum_n (y_n - \bar{w}^T\bar{x}_n)^2 + \frac{\sigma^2}{\sigma_0^2} \cdot \bar{w}^T\bar{w} \xrightarrow[\bar{w}]{} \min}$$

$\frac{\sigma^2}{\sigma_0^2} = \alpha$

Ridge regression

L_2-regularization

regularizer

$\alpha \cdot \bar{w}^T\bar{w}$

$\cancel{(A\bar{x})}$ $\cancel{\det A = 0}$

$\underline{(A + \lambda I)} \rightsquigarrow \ldots (\lambda) \ldots$

$\rightarrow 0$

$$(\bar y - X\bar w)^T(\bar y - X\bar w) + \alpha \cdot \bar w^T \bar w \to \min$$

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

$$\bar y^T \bar y - 2\bar w^T(X^T\bar y) + \bar w^T(X^T X + \alpha I)\bar w \to \min$$

$$\boxed{\bar w_* = (X^T X + \alpha I)^{-1} X^T \bar y}$$

Lasso regression

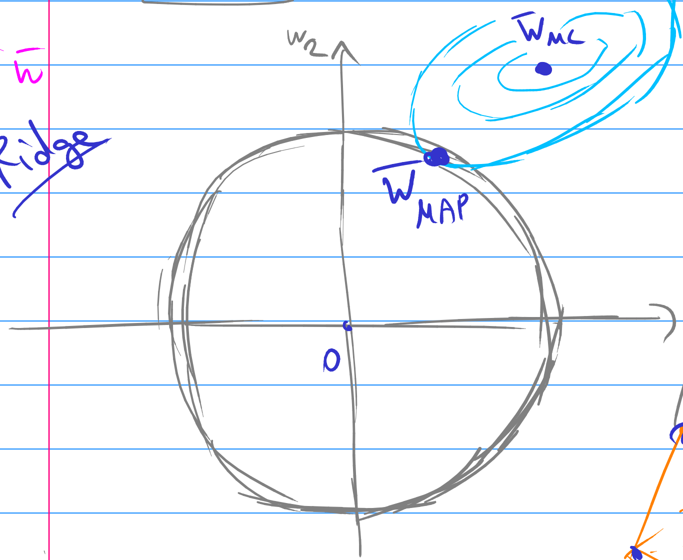
L_1-regularization

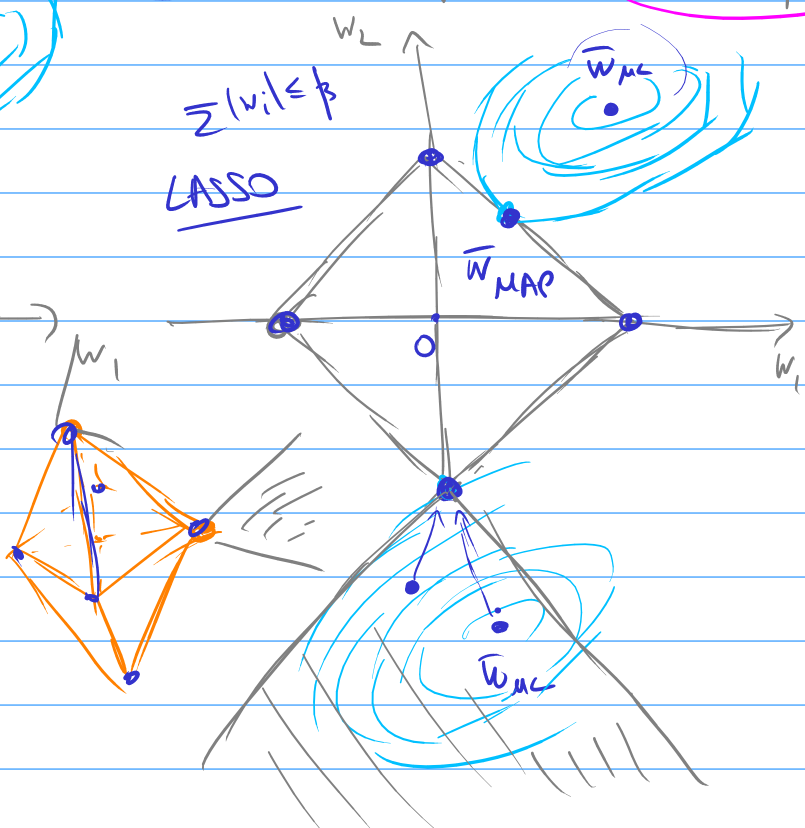
$$\sum_n (y_n - \bar w^T \bar x_n)^2 + \alpha \sum_{i=1}^{d} |w_i| \xrightarrow[\bar w]{} \min$$

$f(\bar w) + \underline{\alpha R(\bar w)} \to \min \iff \boxed{f(\bar w) \to \min}$ при усл. $R(\bar w) \le \beta$

$\bar w^T \bar w$

~~Ridge~~



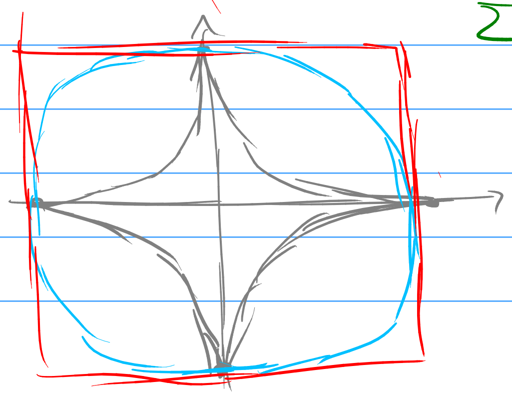


Bridge regression L_q-regularization

$\Sigma |w_i|^q$

$q < 1$

$q = 2$



Elastic Net

$$\Sigma(y_n - \bar w^T \bar x_n)^2 + \alpha \Sigma w_i^2 + \beta \Sigma |w_i|$$

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

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

$$p(\bar w) = \mathcal{N}(\bar w|\bar\mu_0,\Sigma_0) = \frac{1}{(2\pi)^{d/2}\sqrt{\det\Sigma_0}}\, e^{-\frac12(\bar w-\bar\mu_0)^T\Sigma_0^{-1}(\bar w-\bar\mu_0)}$$

$$\ln p(\bar w|D) = \text{Const} - \frac12(\bar w-\bar\mu)^T\Sigma^{-1}(\bar w-\bar\mu)$$

$$\ln p(\bar w|D) = \text{Const} - \frac{d}{2}\ln 2\pi - \frac12\ln\det\Sigma_0 - \frac12(\bar w-\bar\mu_0)^T\Sigma_0^{-1}(\bar w-\bar\mu_0) - \frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(y_n-\bar w^T\bar x_n)^2 =$$

$$= \text{Const} - \frac12(\bar w-\bar\mu_0)^T\Sigma_0^{-1}(\bar w-\bar\mu_0) - \frac{1}{2\sigma^2}(\bar y - X\bar w)^T(\bar y - X\bar w)$$

$$= \text{Const} - \frac12\bar w^T\Sigma_0^{-1}\bar w + \cancel{\tfrac12}\,\bar w^T\Sigma_0^{-1}\bar\mu_0 + \cancel{\tfrac12\bar\mu_0^T\Sigma_0^{-1}\bar w} - \cancel{\tfrac12\bar\mu_0^T\Sigma_0^{-1}\bar\mu_0} - \cancel{\tfrac{1}{2\sigma^2}\bar y^T\bar y} + \cancel{\tfrac{1}{2\sigma^2}\bar y^TX\bar w} + \frac{1}{\cancel{2}\sigma^2}\bar w^TX^T\bar y - \frac{1}{2\sigma^2}\bar w^TX^TX\bar w =$$

$$= \text{Const} - \frac12\bar w^T\left(\Sigma_0^{-1} + \frac{1}{\sigma^2}X^TX\right)\bar w + \bar w^T\left(\Sigma_0^{-1}\bar\mu_0 + \frac{1}{\sigma^2}X^T\bar y\right)$$

$$\ln p(\bar w|D) = \text{Const} - \frac12\bar w^T\Sigma^{-1}\bar w + \bar w^T\Sigma^{-1}\bar\mu$$

$$\bar w_{MAP} = \bar\mu$$

$$\begin{cases}\Sigma^{-1} = \Sigma_0^{-1} + \dfrac{1}{\sigma^2}X^TX \\[2ex] \bar\mu = \Sigma\left(\Sigma_0^{-1}\bar\mu_0 + \dfrac{1}{\sigma^2}X^T\bar y\right)\end{cases}$$

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

$$(\bar\mu_0,\Sigma_0) \xrightarrow{D} (\bar\mu,\Sigma) \xrightarrow{D'} (\bar\mu',\Sigma') \rightsquigarrow \ldots$$

$$\mathbb{E}_{\bar w\sim p(\bar w|D)}\left[p(y|\bar x,\bar w)\right]$$

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

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

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

$$= \int_{\mathbb{R}^d} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\bar{w}^T\bar{x})^2} \cdot \frac{1}{(2\pi)^{d/2}(\det\Sigma)^{1/2}} e^{-\frac{1}{2}(\bar{w}-\bar{\mu})^T\Sigma^{-1}(\bar{w}-\bar{\mu})} \, d\bar{w}$$

$$\boxed{\operatorname{erf}(a) = \int_{-\infty}^{a} \mathcal{N}(x|0,1)\,dx}$$

$$\int_{\mathbb{R}^d} \frac{1}{(2\pi)^{d/2}\det\Sigma'} e^{-\frac{1}{2}(\bar{w}-\bar{\mu}')^T\Sigma'^{-1}(\bar{w}-\bar{\mu})} \, d\bar{w} = 1$$

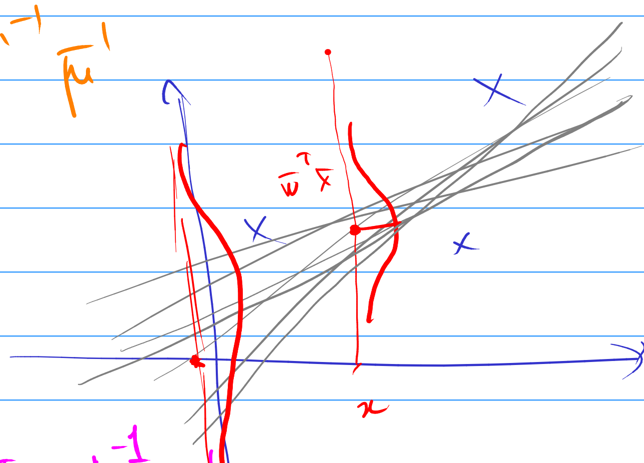
$$-\frac{1}{2}\ln 2\pi\sigma^2 - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln\det\Sigma - \frac{1}{2\sigma^2}\left(y^2 - 2y(\bar{w}^T\bar{x}) + \bar{w}^T(\bar{x}\bar{x}^T)\bar{w}\right) -$$

$$-\frac{1}{2}\left(\bar{w}^T\Sigma^{-1}\bar{w} - 2\bar{w}^T\Sigma^{-1}\bar{\mu} + \bar{\mu}^T\Sigma^{-1}\bar{\mu}\right) =$$

$$= \text{Const} - \frac{1}{2}\bar{w}^T\Big(\underbrace{\frac{1}{\sigma^2}\bar{x}\bar{x}^T + \Sigma^{-1}}_{\text{"}\Sigma'^{-1}}\Big)\bar{w} + \bar{w}^T\Big(\underbrace{\frac{y}{\sigma^2}\bar{x} + \Sigma^{-1}\bar{\mu}}_{\Sigma'^{-1}\bar{\mu}'}\Big) - \frac{y^2}{2\sigma^2}$$

$$\boxed{-\frac{1}{2}\bar{\mu}'^T\Sigma'^{-1}\bar{\mu}'} + \frac{1}{2}\bar{\mu}'^T\Sigma'^{-1}\bar{\mu}'$$

$$\int e^{-\frac{1}{2}(\ldots)}\,d\bar{w} = (2\pi)^{d/2}(\det\Sigma')^{1/2}$$



$$\ln p(y|\bar{x},D) = \text{Const} - \frac{y^2}{2\sigma^2} + \frac{1}{2}\bar{\mu}'^T\Sigma'^{-1}\bar{\mu}$$

$$\bar{\mu}' = \Sigma'\left(\frac{y}{\sigma^2}\bar{x} + \Sigma^{-1}\bar{\mu}\right)$$

$$\bar{\mu}'^T\Sigma'^{-1}\bar{\mu}' = (\ldots)^T\cdot\cancel{(\Sigma')^T}\cdot\cancel{(\Sigma')^{-1}}\cdot\Sigma'\cdot(\ldots) =$$

$$= \left(y/\sigma^2\cdot\bar{x} + \Sigma^{-1}\bar{\mu}\right)^T\left(\Sigma^{-1} + \frac{1}{\sigma^2}\bar{x}\bar{x}^T\right)^{-1}\left(\frac{y}{\sigma^2}\bar{x} + \Sigma^{-1}\bar{\mu}\right)$$

$$= \left(\frac{y}{\sigma^2}\bar{x} + \Sigma^{-1}\bar{\mu}\right)^T \Sigma' \left(\frac{y}{\sigma^2}\bar{x} + \Sigma^{-1}\bar{\mu}\right)$$

$$y \cdot \left(\frac{1}{\sigma^2}\bar{x}^T \Sigma' \Sigma^{-1}\bar{\mu}\right)$$

К-нт при y^2: $-\frac{1}{2}\left(\frac{1}{\sigma^2} - \frac{1}{\sigma^4}\bar{x}^T \boxed{\Sigma' \bar{x}}\right)$

$\times\Sigma'$

$$\left(\Sigma^{-1} + \frac{1}{\sigma^2}\bar{x}\bar{x}^T\right)\Sigma'\bar{x} = \bar{x} + \frac{1}{\sigma^2}\bar{x}(\bar{x}^T\Sigma\bar{x}) = \left(\frac{1}{\sigma^2}\bar{x}^T\Sigma\bar{x} + 1\right)\bar{x}$$

$=$

$\times\Sigma'$ $(\Sigma'^{-1}\Sigma)\bar{x}$

$$\Sigma\bar{x} = \left(\frac{1}{\sigma^2}\bar{x}^T\Sigma\bar{x} + 1\right)\Sigma'\bar{x}$$

К-нт при y^2: $-\frac{1}{2}\left(\frac{1}{\sigma^2} - \frac{1}{\sigma^4}\frac{1}{\frac{1}{\sigma^2}\bar{x}^T\Sigma\bar{x}+1}\bar{x}^T\Sigma\bar{x}\right) =$

$$= -\frac{1}{2\sigma^2}\left(1 - \frac{\bar{x}^T\Sigma\bar{x}}{\bar{x}^T\Sigma\bar{x}+\sigma^2}\right) = -\frac{1}{2(\bar{x}^T\Sigma\bar{x}+\sigma^2)}$$

$$\boxed{\sigma_{pred}(\bar{x})^2 = \sigma^2 + \bar{x}^T\Sigma\bar{x}}$$

$$e^{-\frac{1}{2\sigma^2_{pred}}(y - \mu_{pred})^2} \qquad \frac{\mu_{pred}}{\sigma^2_{pred}}$$

К-нт при y: $\frac{1}{\sigma^2}\bar{x}^T\Sigma'\Sigma^{-1}\bar{\mu}$

$$\mu_{pred}(\bar{x}) = (\sigma^2 + \bar{x}^T\Sigma\bar{x}) \cdot \frac{1}{\sigma^2}\bar{x}^T(\Sigma'\Sigma^{-1}\bar{\mu}) =$$

$$= \frac{1}{\sigma^2}(\sigma^2 + \bar{x}^T\Sigma\bar{x})\,\bar{\mu}^T\Sigma^{-1}(\Sigma'\bar{x}) = \cancel{\frac{1}{\sigma^2}(\sigma^2 + \bar{x}^T\Sigma\bar{x})}\,\bar{\mu}^T\cancel{\Sigma^{-1}} \cdot \frac{\cancel{\Sigma}\bar{x}}{\cancel{\frac{1}{\sigma^2}\bar{x}^T\Sigma\bar{x}+1}}$$

$$\boxed{\mu_{pred}(\bar{x}) = \bar{\mu}^T\bar{x}}$$