$$\sum_{n=1}^{N}\left(y_n-\bar w^T\bar x_n\right)^2+\frac{\lambda}{2}\cdot\|\bar w\|_2^2 \xrightarrow[\bar w]{} \min$$

$\|\bar w\|_2^2 = \sum w_i^2$ — regularizer

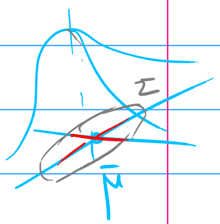
$$(\bar y-X\bar w)^T(\bar y-X\bar w)+\underline{\frac{\lambda}{2}\bar w^T\bar w} \to \min$$

$$-X^T\bar y+(X^TX+\lambda I)\bar w=0 \quad \overset{\lambda \bar w}{=}$$

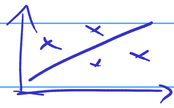
$$\bar w^*=(X^TX+\lambda I)^{-1}X^T\bar y$$

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

$$\frac{\lambda}{2}\bar w^T\bar w = \text{const} - \ln e^{-\frac{\lambda}{2}\bar w^T\bar w} = \text{const} + \ln \mathcal{N}\left(\bar w \mid \bar 0, \frac{1}{\lambda}I\right)$$



$$p(\bar w\mid D) \propto \underline{p(\bar w)\,p(D\mid\bar w)} = \prod_n \mathcal{N}(y_n\mid \bar w^T\bar x_n,\sigma^2) \xrightarrow[\bar w]{} \max$$



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

$$y\sim\bar w^T\bar x$$

$$p(D\mid\bar w)\xrightarrow[\bar w]{}\max$$

$$\bar w_{ML}=(X^TX)^{-1}X^T\bar y$$

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

$$\ln p(\bar w\mid D)=\text{Const}+\ln p(D\mid\bar w)+\ln p(\bar w)=$$

$$=\text{Const}-\cancel{\frac{N}{2}\ln(2\pi\sigma^2)}-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(y_n-\bar w^T\bar x_n)^2-\cancel{\frac{d}{2}\ln 2\pi}-\cancel{\frac{1}{2}\ln\det\Sigma}-$$
$$-\frac{1}{2}(\bar w-\bar\mu)^T\Sigma^{-1}(\bar w-\bar\mu)$$

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

$\bar w^T\Sigma^{-1}\bar\mu$

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

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

$\mu',\ \Sigma'$

$$\ln p(\bar w\mid D)=\text{Const}-\frac{1}{2}(\bar w-\bar\mu')^T\Sigma'^{-1}(\bar w-\bar\mu')=\text{Const}-\frac{1}{2}\bar w^T\Sigma'^{-1}\bar w+\bar w^T(\Sigma'^{-1}\bar\mu')$$

$$\begin{cases}\Sigma'^{-1}=\Sigma^{-1}+\frac{1}{\sigma^2}X^TX\\[2ex] \bar\mu'=\Sigma'\cdot\left(\Sigma^{-1}\bar\mu+\frac{1}{\sigma^2}X^T\bar y\right)\end{cases}$$

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

$D=\{X,\bar y\}$

$D'=\{X',\bar y'\}$

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

$$\Sigma'=\frac{1}{\sigma_0^2}I+\frac{1}{\sigma^2}X^TX$$

$$\bar w_{MAP}=\arg\max p(\bar w\mid D)=\bar\mu'$$

$$\bar\mu'=\left(\lambda I+\frac{1}{\sigma^2}X^TX\right)^{-1}\frac{1}{\sigma^2}X^T\bar y$$

$$=\left(\frac{\sigma^2}{\sigma_0^2}I+X^TX\right)^{-1}\cdot X^T\bar y$$

λ

$p(\bar{w}|D) \propto \underbrace{p(\bar{w})}_{} \, \underbrace{p(D|\bar{w})}_{}$ $\quad \leftarrow N(\bar{w}|0, \sigma_0^2 I)$

$\rightsquigarrow$ ridge

$\sum(\_)^2 + \frac{\lambda}{2}\|w\|_2^2 \to \min$

lasso

$\sum(\_)^2 + \frac{\lambda}{2}\sum|w_i| \to \min$

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

$\sum|w_i| \le a$

$$\sum_n (y_n - \bar{w}^T\bar{x}_n)^2 + \frac{\lambda}{2}\|\bar{w}\|_2^2 \to \min$$

$$\Updownarrow$$

$$\sum_n (y_n - \bar{w}^T\bar{x}_n)^2 \xrightarrow[\bar{w}]{} \min, \quad \|\bar{w}\|_2^2 \le a$$