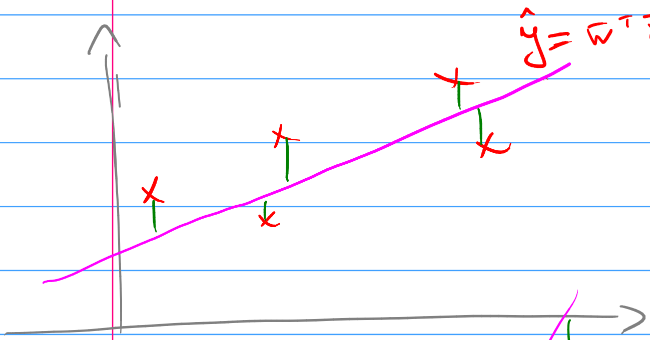
$$\underbrace{p(\theta|D)}_{\text{posterior}} \propto \overbrace{p(\theta)}^{\text{prior}}\,\overbrace{p(D|\theta)}^{\text{likelihood}}$$

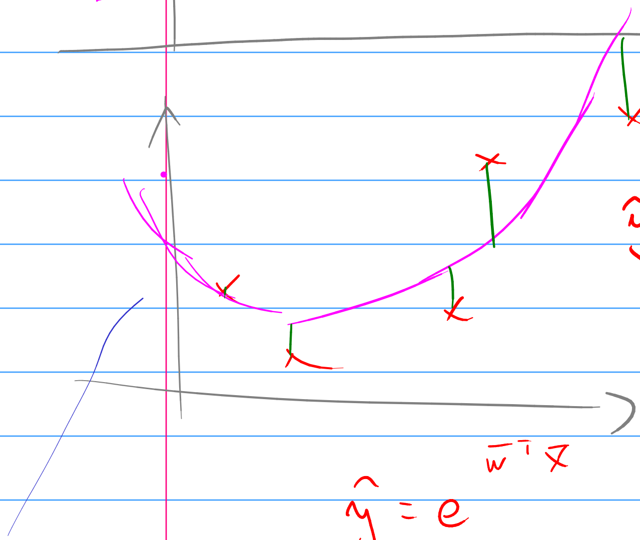
$$p(x|D) = \int \underbrace{p(\theta|D)}_{\text{posterior}}\,\overbrace{p(x|\theta)}^{\text{likelihood}}\,d\theta$$




$$p(D|\bar{w}) = \prod_{(\bar{x},y)\in D} p(y|\bar{w},\bar{x}) =$$

$$= \prod_{(\bar{x},y)} \mathcal{N}(y\,|\,\bar{w}^T\bar{x}, \sigma^2) \xrightarrow[\bar{w}]{} \max$$

$$\Updownarrow$$



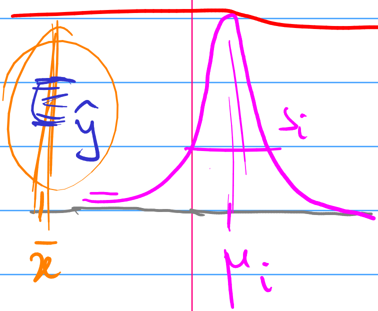
$$\hat{y} = w_0 + w_1 x + w_2 x^2 + \ldots \qquad \sum_{(\bar{x},y)\in D} (y - \bar{w}^T\bar{x})^2 \xrightarrow[\bar{w}]{} \min$$

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

$$\hat{y} = \bar{w}^T \varphi(\bar{x})$$

features

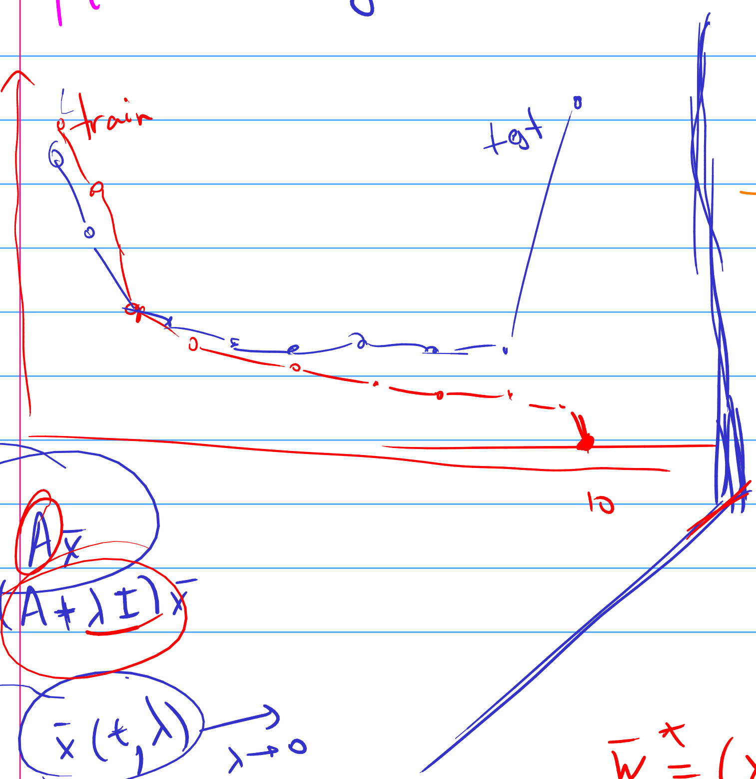
$$\hat{y} = e^{\bar{w}^T\bar{x}} \qquad \ln y = \bar{w}^T\bar{x} + \varepsilon$$




$$\varphi_i(x) = e^{-\frac{1}{2s_i^2}(x-\mu_i)^2}$$

$$\bar{\varphi}(x) = \begin{pmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^k \end{pmatrix}$$

$$\hat{y}(x) = \sum w_i \varphi_i(x)$$




$$L(\bar{w}) = \sum_n (y_n - \bar{x}_n^T\bar{w})^2 + \boxed{\lambda \cdot \sum w_i^2}$$

$$\cdot \sum |w_i|$$

ridge regression

regularizer

$$\bar{x} = A\bar{x}$$

$$\bar{x} = (A + \lambda I)\bar{x}$$

$$\bar{x}(t,\lambda) \xrightarrow[\lambda \to 0]{}$$

$$(\bar{y} - X\bar{w})^T(\bar{y} - X\bar{w}) + \lambda \bar{w}^T\bar{w}$$

$$-2X^T\bar{y} + 2X^TX\bar{w} + 2\lambda\bar{w} = 0$$

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

$$p(\bar{w}|D) \propto p(\bar{w}) \prod_n p(y_n|\bar{w},\bar{x}_n) =$$

$$= \mathcal{N}(\bar{w}|\bar{0},\alpha I) \prod_n \mathcal{N}(y_n|\bar{w}^T\bar{x}_n,\sigma^2) \xrightarrow[\bar{w}]{} \max$$

$$\ln p(\bar{w}|D) = \text{const} - \frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_n (y_n - \bar{w}^T\bar{x}_n)^2 -$$

$$- \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln\det(\alpha I) - \frac{1}{2\alpha}\bar{w}^T\bar{w} = \circledast$$

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

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

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

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

var. $w_i$

$$\ln\mathcal{N}(\bar{x}|\bar{\mu},\Sigma) = -\frac{d}{2}\ln 2\pi - \frac{1}{2}\ln\det\Sigma - \frac{1}{2}(\bar{x}-\bar{\mu})^T\Sigma^{-1}(\bar{x}-\bar{\mu}) =$$

$$= \text{const} - \frac{1}{2}(\bar{x}^T\Sigma^{-1}\bar{x} - 2\bar{x}^T\Sigma^{-1}\bar{\mu} + \bar{\mu}^T\Sigma^{-1}\bar{\mu})$$

$$\circledast = \text{const} - \frac{1}{2\sigma^2}(\bar{y} - X\bar{w})^T(\bar{y} - X\bar{w}) - \frac{1}{2\alpha}\bar{w}^T\bar{w} =$$

$$= \text{const} - \frac{1}{2\sigma^2}(\bar{y}^T\bar{y} - 2\bar{w}^TX^T\bar{y} + \bar{w}^TX^TX\bar{w}) - \frac{1}{2\alpha}\bar{w}^T\bar{w}$$

$$-\frac{1}{2}\left(-\frac{1}{\sigma^2}\cdot 2\bar{w}^TX^T\bar{y}\right)$$

$$-\frac{1}{2\sigma^2}\bar{w}^T X^TX\bar{w} - \frac{1}{2\alpha}\bar{w}^T I\bar{w}$$

$$\Sigma'^{-1}\bar{\mu}' = \frac{1}{\sigma^2}X^T\bar{y}$$

$$= -\frac{1}{2}\bar{w}^T\underbrace{\left(\frac{1}{\sigma^2}X^TX + \frac{1}{\alpha}I\right)}_{=\Sigma'^{-1}}\bar{w}$$

$$p(\bar w) = \mathcal{N}(\bar w \mid \bar\mu_0, \Sigma_0)$$

$$\ln p(\bar w \mid D) = const - \frac{1}{2\sigma^2}(\bar y - X\bar w)^T(\bar y - X\bar w) - \frac{1}{2}(\bar w - \bar\mu_0)^T\Sigma_0^{-1}(\bar w - \bar\mu_0)$$

$$= const - \frac{1}{2}\Big(\cancel{\frac{1}{\sigma^2}\bar y^T\bar y} - 2\frac{1}{\sigma^2}\bar w^T X^T\bar y + \frac{1}{\sigma^2}\bar w^T X^T X\bar w$$

$$+ \bar w^T\Sigma_0^{-1}\bar w - 2\bar w^T\Sigma_0^{-1}\bar\mu_0 + \cancel{\bar\mu_0^T\Sigma_0^{-1}\bar\mu_0}\Big)$$

$$= const - \frac{1}{2}\Big(\bar w^T\overbrace{\Big(\Sigma_0^{-1} + \frac{1}{\sigma^2}X^TX\Big)}^{=\Sigma^{-1}}\bar w -$$

$$- 2\bar w^T\underbrace{\Big(\Sigma_0^{-1}\bar\mu_0 + \frac{1}{\sigma^2}X^T\bar y\Big)}_{=\Sigma^{-1}\bar\mu}\Big)$$

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

$$\begin{cases} \Sigma^{-1} = \Sigma_0^{-1} + \frac{1}{\sigma^2}X^TX \\ \bar\mu = \Sigma\cdot\Big(\Sigma_0^{-1}\bar\mu_0 + \frac{1}{\sigma^2}X^T\bar y\Big) \end{cases}$$

$\mu_0 = 0$, $\Sigma_0 = \alpha I$

$$\Sigma^{-1} = \frac{1}{\sigma^2}X^TX + \frac{1}{\alpha}I$$

$$\bar\mu = \Sigma\cdot\frac{1}{\sigma^2}X^T\bar y$$

Lasso regression

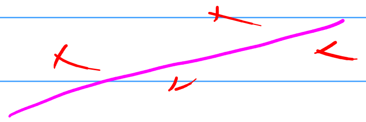
$$L(\bar w) = \sum_n (y_n - \bar x_n^T\bar w)^2 + \lambda\cdot\sum_{i=1}^{d}|w_i|$$

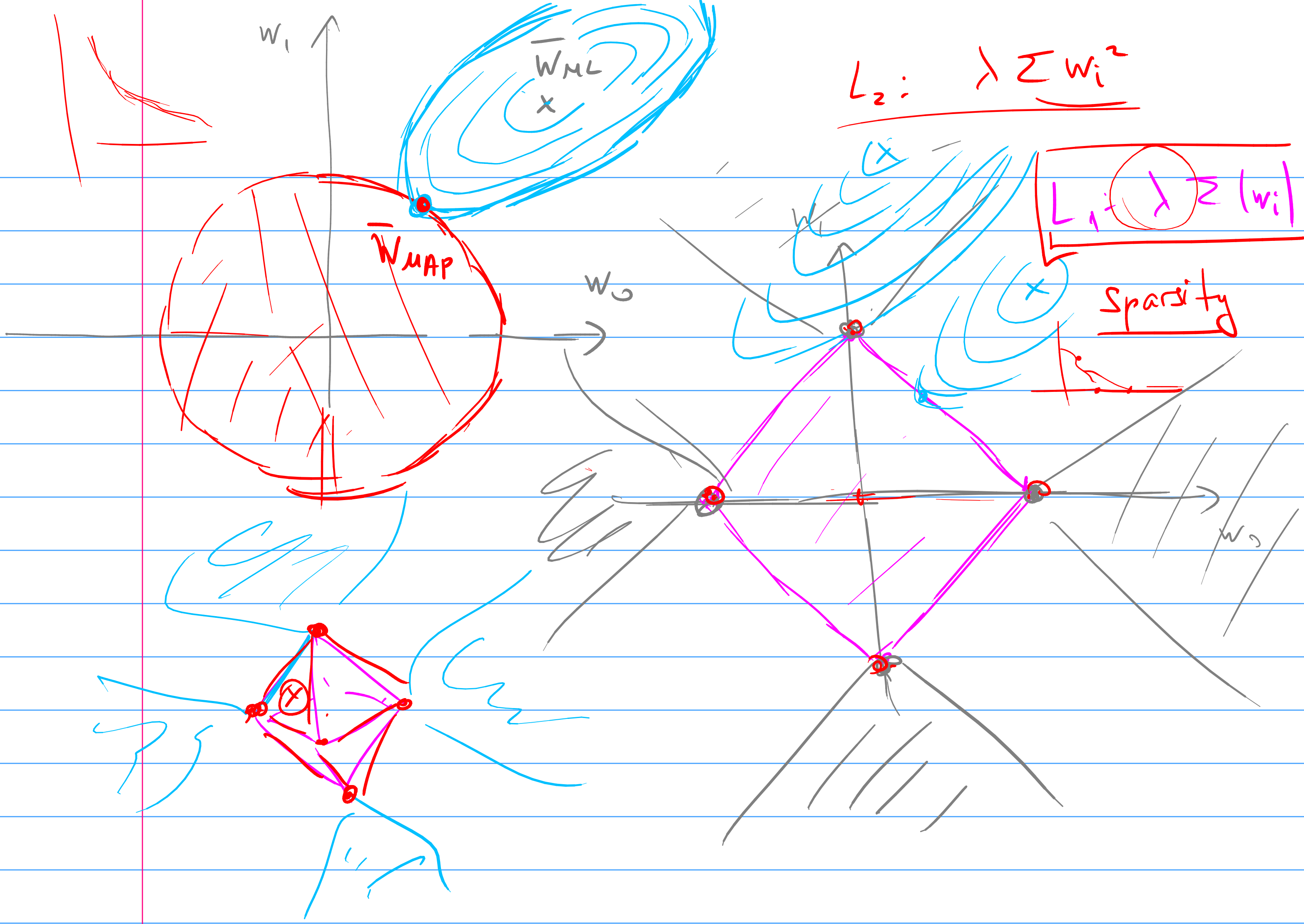
$$p(\bar w) \propto e^{-\frac{\lambda}{2}\sum_{i=1}^{d}|w_i|}$$

$\sum_n (y_n - \bar x_n^T\bar w)^2 \to \min$

при условии $\sum|w_i| \le \dots$

$\hat y = w_0 + w_1 x$






$$p(D|\bar{w}) \to \max$$

$$p(\bar{w}|D) \to \max$$

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

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

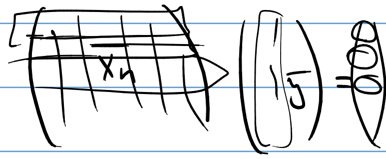
$$= \int \frac{1}{\sqrt{2\pi\sigma^2}} \, \frac{1}{(2\pi)^{d/2}\sqrt{\det \Sigma}} \cdot e^{-\frac{1}{2\sigma^2}(y - \bar{x}^T\bar{w})^2 - \frac{1}{2}(\bar{w}-\bar{\mu})^T \Sigma^{-1} (\bar{w}-\bar{\mu})} \, d\bar{w}$$

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

$$p(y|\bar{x}, D) = \mathcal{N}(y \mid \bar{w}_{MAP}^T \bar{x}, \sigma_N^2)$$

$$'' \; \sigma^2 + \bar{x}^T \Sigma_N \bar{x}$$

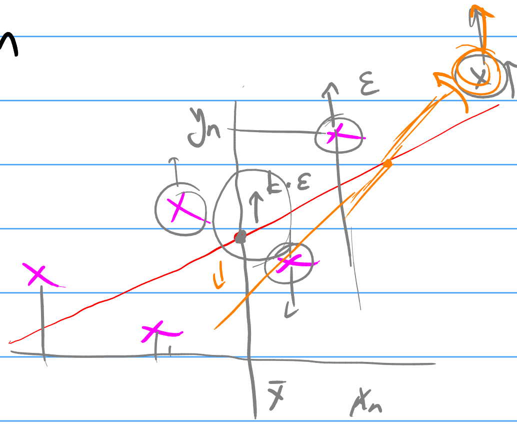
$$\hat{y} = \bar{x}^T \left( \frac{1}{\sigma^2} \Sigma_N X^T \bar{y} \right) = \frac{1}{\sigma^2} \bar{x}^T \cdot \Sigma_N X^T \bar{y} =$$




$$= \frac{1}{\sigma^2} \bar{x}^T \Sigma_N \sum_{n=1}^{N} \bar{x}_n y_n =$$

$$X^T \bar{y} = \sum_n \bar{x}_n y_n$$

$$= \sum_{n=1}^{N} \frac{1}{\sigma^2} \left( \bar{x}^T \Sigma_N \bar{x}_n \right) \cdot y_n$$




$$\hat{y} = \sum_{n=1}^{N} k(\bar{x}, \bar{x}_n)\, y_n$$