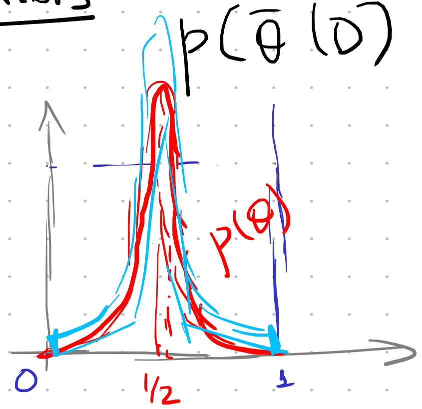
① Priors

$$p(\bar{\theta} \mid D) = \frac{p(\bar{\theta})\, p(D \mid \bar{\theta})}{p(D)}$$

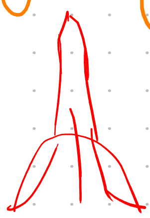


$p(\bar{\theta}) = \text{Unif}([0,1])$

$p(\text{heads} \mid D) = \frac{n+1}{n+m+2}$

$$p(\theta) \quad \times \quad p(D \mid \theta) \quad \propto \quad p(\theta \mid D)$$

$$\theta^{\alpha-1}(1-\theta)^{\beta-1} \times \theta^{n}(1-\theta)^{m} \propto \theta^{\alpha+n-1}(1-\theta)^{\beta+m-1}$$



$$p(\theta) = \text{Beta}(\theta \mid \alpha, \beta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$D = \{n, m\}$

$$\underline{\text{Beta}(\theta; \alpha, \beta)} \times \underline{\theta^{n}(1-\theta)^{m}} \propto \underline{\text{Beta}(\theta; \alpha+n, \beta+m)}$$

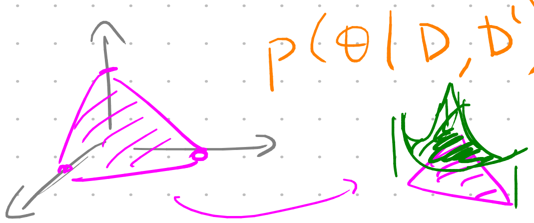
equivalent sample size

conjugate priors

$D' = \{n', m'\}$

$$\text{— — —} \times \theta^{n'}(1-\theta)^{m'} \propto \text{Beta}(\theta; \alpha+n+n', \beta+m+m')$$

$p(\theta \mid D, D')$



$$\underline{p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}}$$

$$p(D \mid \bar{\theta}) = \underline{\theta_1^{n_1}\theta_2^{n_2}\ldots\theta_k^{n_k}}$$

$p(k \mid D) = \frac{n_k+\alpha_k+1}{N+\Sigma\alpha+k}$

$\bar{\theta} = \{\theta_1, \ldots, \theta_k\}$

$\Sigma\theta_k = 1$

$D = \{n_1, n_2, \ldots, n_k\}$

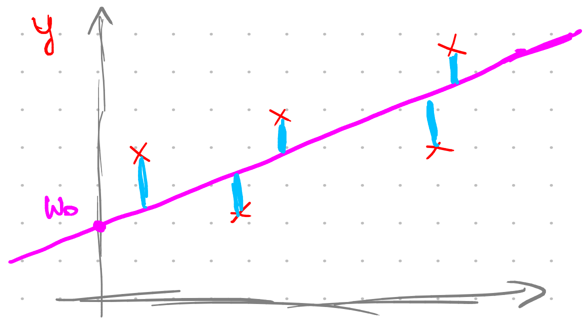
$$p(\bar{\theta}) \propto \underline{\theta_1^{\alpha_1-1}\ldots\theta_k^{\alpha_k-1}}$$

$$p(\bar{\theta}) = \frac{1}{\text{Dir}(\bar{\alpha})}\theta_1^{\alpha_1-1}\ldots\theta_k^{\alpha_k-1}$$

$$p(\bar{\theta}) = \text{Dir}\left(\bar{\theta}; \frac{1}{a}, \ldots, \frac{1}{a}\right)$$

sparsity

② Linear regression



$$D = \{(\bar x_n, y_n)\}_{n=1}^{N}$$

$$y \sim w_0 + w_1 x_1 + \dots + w_d x_d$$

$$\bar x \mapsto \begin{pmatrix} 1 \\ \bar x \end{pmatrix}$$

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

$$\frac{1}{N}\sum_{n=1}^{N} |y_n - \hat y_n| = \frac{1}{N}\sum_n |y_n - \bar w^T \bar x_n| \xrightarrow{\bar w} \min$$

$$\sum (y_n - \bar w^T \bar x_n)^2 \xrightarrow{\bar w} \min$$

$$d(y_n, \bar w^T \bar x_n) \qquad f(|y_n - \bar w^T \bar x_n|)$$

$$\sum (\ )^4 \to \dots \qquad \sum e^{|\ |} \qquad \sum \dots$$

$$\sum_{n=1}^{N} (y_n - \bar w^T \bar x_n)^2 \xrightarrow{\bar w} \min$$

$$(\bar y - X\bar w)^T (\bar y - X\bar w) \xrightarrow{\bar w} \min$$

$$\bar y^T \bar y - (X\bar w)^T \bar y - \bar y^T (X\bar w) + (X\bar w)^T (X\bar w) =$$

$$= Const - 2\bar w^T X^T \bar y + \bar w^T X^T X \bar w \xrightarrow{\bar w} \min$$

$$\nabla_{\bar x} (\bar x^T \bar a) = \nabla_{\bar x}(a_1 x_1 + \dots + a_d x_d) = \bar a$$

$$\nabla_{\bar x} (\bar x^T \bar x) = \nabla_{\bar x}(x_1^2 + \dots + x_d^2) = 2\bar x$$

$$\nabla_{\bar x} (\bar x^T A \bar x) = \nabla_{\bar x}\Big(\sum_{i,j} a_{ij} x_i x_j\Big) = A\bar x + A^T \bar x = (A + A^T)\bar x$$

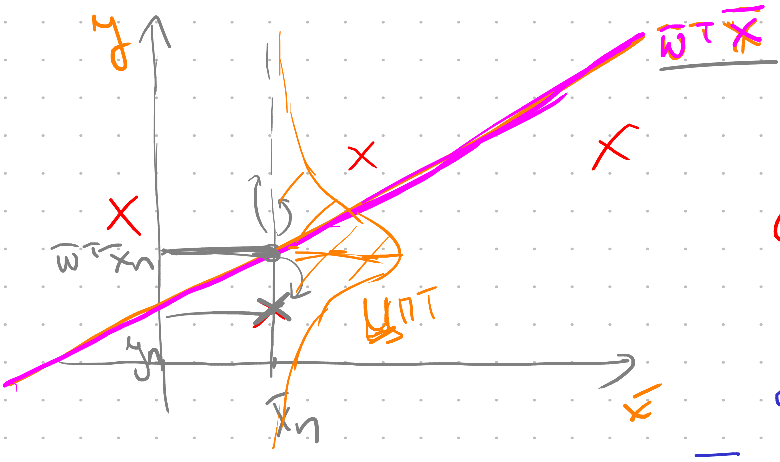
$$\frac{\partial(\sum_{i,j} a_{ij} x_i x_j)}{\partial x_k} = \frac{\partial(\sum_{i\neq k} a_{ik} x_i x_k + \sum_{j \neq k} a_{kj} x_k x_j + a_{kk} x_k^2)}{\partial x_k} =$$

$$= \sum_{i \neq k} a_{ik} x_i + \sum_{j\neq k} a_{kj} x_j + 2 a_{kk} x_k = \sum a_{ik} x_i + \sum a_{kj} x_j$$

$$\nabla_{\bar w} L(\bar w) = -2X^T\bar y + 2(X^TX)\cdot\bar w = 0$$

$$(X^TX)\bar w = X^T\bar y \qquad \boxed{\bar w_* = (X^TX)^{-1}X^T\bar y}$$

$$D = \{(\bar x_n, y_n)\}_{n=1}^N$$



$$p(D|\bar w) \overset{1}{=}$$

$$\overset{1}{=} \prod_{n=1}^N p(y_n|\underbrace{\bar x_n, \bar w}) =$$

$$= \prod_{n=1}^N \mathcal{N}(y_n | \bar w^T\bar x_n, \sigma^2) =$$

$$= \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_n-\bar w^T\bar x_n)^2} \xrightarrow{\bar w} \max \qquad \bar w_{ML}$$

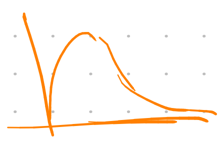
$$\log p(D|\bar w) = -\frac{N}{2}\log(2\pi\sigma^2) - \sum_{n=1}^N \frac{1}{2\sigma^2}(y_n - \bar w^T\bar x_n)^2$$

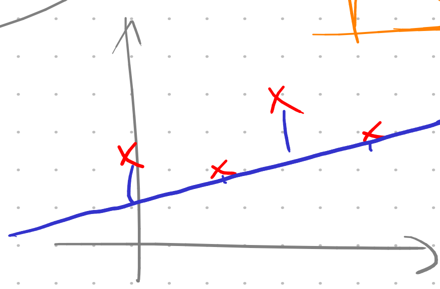
$$= \text{Const} - \frac{1}{2\sigma^2}\boxed{\sum_{n=1}^N (y_n - \bar w^T\bar x_n)^2} \xrightarrow{\bar w} \max$$

$$\xrightarrow{\bar w} \min$$

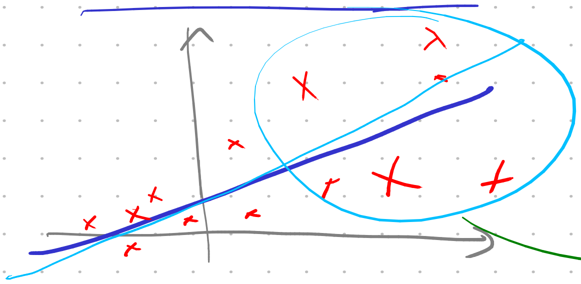
$$p(y_n | \bar x_n^T\bar w) \propto \qquad \sum |y_n - \bar w^T\bar x_n|$$

$$\propto e^{-\text{Const}\cdot|y_n - \bar w^T\bar x_n|}$$



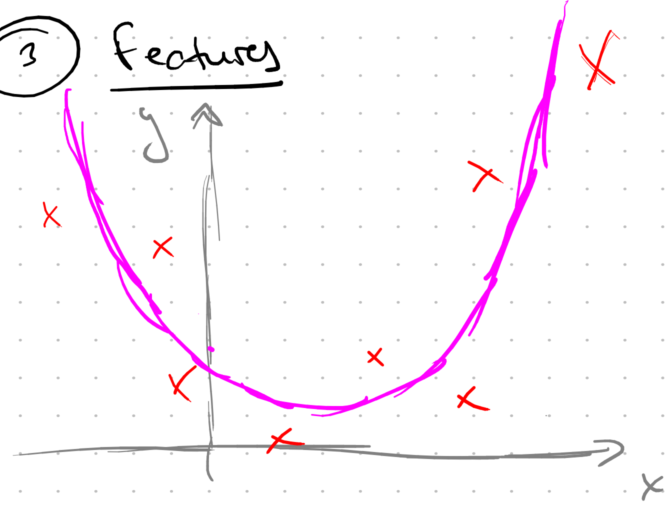


$\boxed{\sigma_n = \sigma}$ – homoscedasticity



$$\sum_n \frac{1}{\sigma_n^2}(y_n - \bar w^T\bar x_n)^2$$

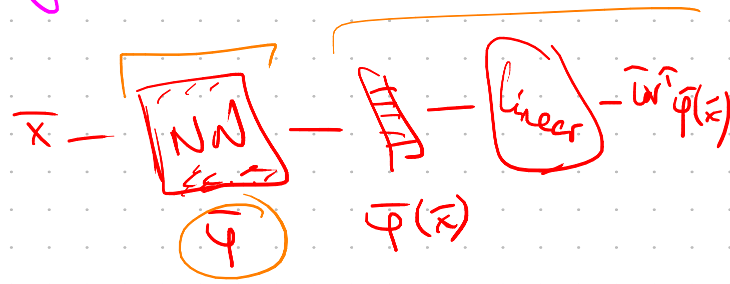
③ Features



$y \sim w_0 + w_1 x + w_2 x^2$

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

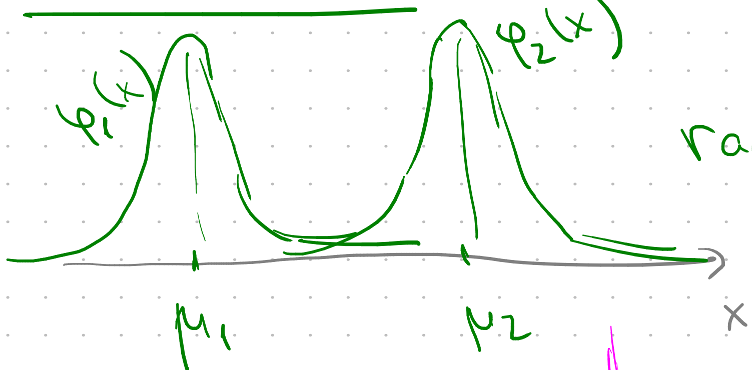
$y \sim \bar{w}^T \bar\varphi(\bar{x})$



polynomial regression

$x \mapsto \begin{pmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^d \end{pmatrix}^T \begin{pmatrix} \bar{w} \end{pmatrix}$

local features



$x - \bar\varphi - \bar\varphi(x) = \begin{pmatrix} \varphi_1(x) \\ \vdots \\ \varphi_d(x) \end{pmatrix}$

radial basis functions

$\varphi_i(x) = c \cdot e^{-c' \cdot (x - \mu_i)^2}$

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