① Coin → die

$$p(D|\theta) = \theta^{n}(1-\theta)^{m}$$

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

Conjugate priors:

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

coin: heads / tails

die: $1,2,\dots,k$ k-sided

✓ $\bar\theta = (\theta_1,\theta_2,\dots,\theta_{k-1})$ $\theta_k = \left(1-\sum_{1}^{k-1}\theta_i\right)$

$\cancel{\bar\theta = (\theta_1,\theta_2,\dots,\theta_k)}$ ~~constraint:~~ $\sum\theta_i = 1$

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

$$p(D|\bar\theta) = \theta_1^{n_1}\,\theta_2^{n_2}\dots\theta_k^{n_k}$$

$$p(\bar\theta|D,\bar\alpha) \propto \theta_1^{\alpha_1+n_1-1}\dots\theta_k^{\alpha_k+n_k-1}$$

×

Conj-prior:

$$p(\bar\theta|\bar\alpha) \propto \theta_1^{\alpha_1-1}\,\theta_2^{\alpha_2-1}\dots\theta_k^{\alpha_k-1}$$

Dirichlet

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

pred. distr.

$$\boxed{p(i|D) = \frac{n_i+1}{\sum n_j + k}}$$

Bayesian smoothing

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

equivalent sample size

uniform + $D = \{\alpha,\beta\}$ → posterior

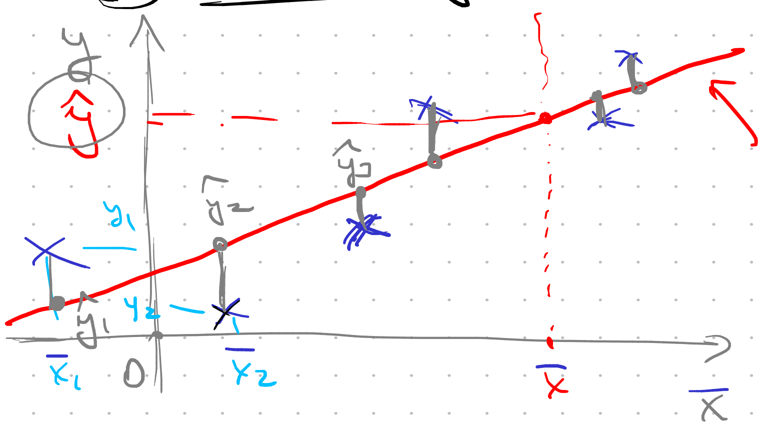
$p(\theta|\cancel{3.5, 4.5})$

$$p\left(\theta\middle|\tfrac{1}{2},\tfrac{1}{2}\right) \propto \frac{1}{\sqrt{\theta(1-\theta)}}$$

Topic models

$$p(\bar\theta|\bar\alpha) = Dir\left(\bar\theta\middle|\tfrac{1}{50},\dots,\tfrac{1}{50}\right)$$

② Linear regression



$$D = \{(\bar{x}_n, y_n)\}_{n=1}^{N}, \quad \bar{x}_n \in \mathbb{R}^d, \quad y_n \in \mathbb{R}$$

$$\hat{y} = w_0 + w_1 x_1 + w_2 x_2 + \dots + w_d x_d$$

$$= \cancel{w_0} + \bar{w}^T \bar{x}$$

$$\bar{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \longmapsto \bar{x}' = \begin{pmatrix} x_0 = 1 \\ x_1 \\ \vdots \\ x_d \end{pmatrix}$$

$$\hat{y} = \bar{w}^T \bar{x}'$$

Mean squared error — MSE

$$L = \sum_{n=1}^{N} (y_n - \bar{w}^T \bar{x}_n)^2 \xrightarrow[\bar{w}]{} \min$$

$$\boxed{\nabla_{\bar{w}} L = 0}$$

$$\sum x_i^2 = \bar{x}^T \bar{x}$$

$$\bar{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \in \mathbb{R}^N, \quad \bar{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_d \end{pmatrix} \in \mathbb{R}^d, \quad X = N \begin{pmatrix} \text{---} \ \bar{x}_1 \ \text{---} \\ \vdots \\ \text{---} \ \bar{x}_n \ \text{---} \\ \vdots \\ \text{---} \ \bar{x}_N \ \text{---} \end{pmatrix}_{d}$$

$N \times d$

$$L = (\bar{y} - X\bar{w})^T (\bar{y} - X\bar{w})$$

$$\nabla_{\bar{x}} f(\bar{x}) = \left(\partial f / \partial x_1 \ \dots \ \partial f / \partial x_d \right)$$

$$\nabla_{\bar{w}} (\bar{w}^T \bar{a}) = \nabla_{\bar{w}} (w_1 a_1 + \dots + w_d a_d) = \bar{a}$$

$$\nabla_{\bar{w}} (\bar{w}^T \bar{w}) = \nabla_{\bar{w}} (w_1^2 + \dots + w_d^2) = 2\bar{w}$$

$$\nabla_{\bar{w}} (\bar{w}^T \underbrace{A}_{d \times d} \bar{w}) = A\bar{w} + A^T \bar{w} = (A + A^T)\bar{w}$$

A – symmetrical $\Rightarrow$ $2A\bar{w}$

$$\bar{w}^T A \bar{w} = \sum_{i,j} a_{ij} w_i w_j$$

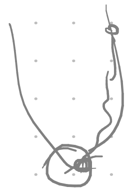
$$\frac{\partial \left(\sum_{i,j} a_{ij} w_i w_j \right)}{\partial w_k} = \frac{\partial \left(a_{kk} w_k^2 + \sum_{i \neq k} a_{ik} w_i w_k + \sum_{j \neq k} a_{kj} w_k w_j \right)}{\partial w_k} =$$

$$= 2a_{kk}w_k + \sum_{i\neq k} a_{ik}w_i + \sum_{j\neq k} a_{kj}w_j = \sum_i a_{ik}w_i + \sum_j a_{kj}w_j$$

$$A^k \quad \left(\bar{w}\right) \qquad (A\bar{w})_k + (A^T\bar{w})_k$$

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

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



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

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

$$\boxed{(X^TX)\bar{w} = X^T\bar{y}}$$

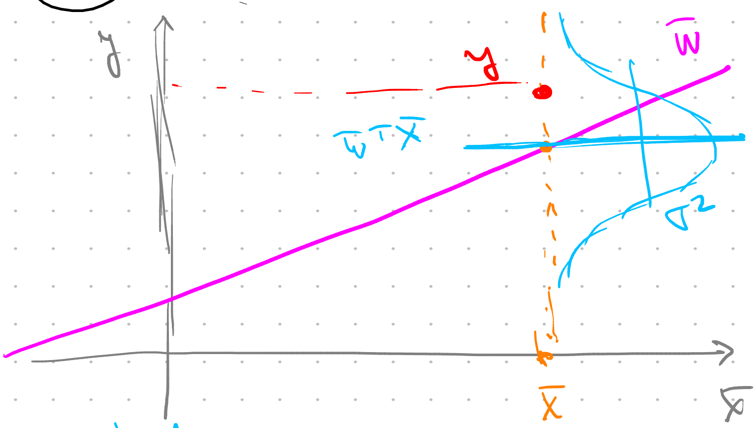
system of equations

$$d_n = y_n - \bar{w}^T\bar{x}_n$$

$$\sum d_n^2 \qquad \sum |d_n| \qquad \sum d_n^4 \qquad \sum e^{|d_n|}$$

③ Assumptions of LR

$$D = \{\bar{x}_n, y_n\}_{n=1}^N$$



Central limit theorem

$$p(y|\bar{x}, \bar{w})$$

$$p(D|\bar{w}) = p(\bar{y}|X, \bar{w})$$

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

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

③ ② ④

$$= \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2\sigma^2}(y_n - \bar{w}^T\bar{x}_n)^2} \xrightarrow[\bar{w}]{\text{Task №1}} \max$$

$$\log p(D|\bar{w}) = \sum_{n=1}^{N}\Big[\underbrace{-\frac{1}{2}\log(2\pi\sigma^2)}_{\text{Const}} - \frac{1}{2\sigma^2}(y_n - \bar{w}^T\bar{x}_n)^2\Big] =$$

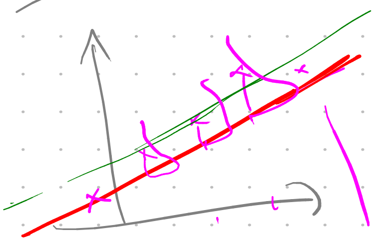
$$= \text{Const} - \frac{1}{2\sigma^2}\boxed{\sum_{n=1}^{N}(y_n - \bar{w}^T\bar{x}_n)^2} \xrightarrow{\bar{w}} \max$$

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

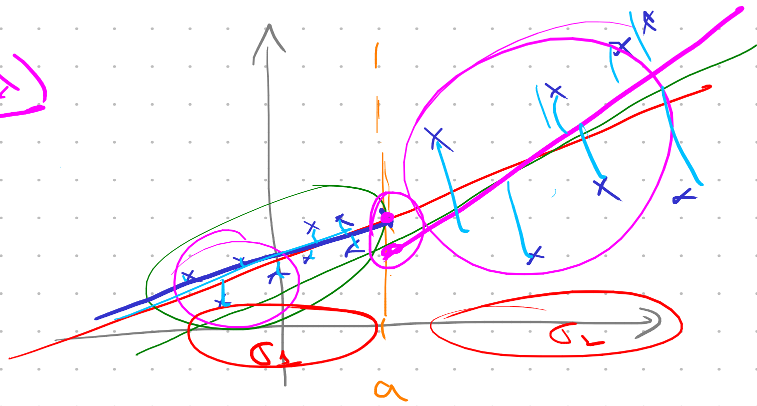
1. y_n are cond-indep. given $\bar{w}, \bar{x}_n$
2. y has a linear dependence on $\bar{x}$
3. $y = \bar{w}^T\bar{x} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$
4. Noise variance is constant, σ^2

$\Rightarrow$ $\sum d_n^2 \to \min$

$\Leftrightarrow$ likelihood $\to \max$



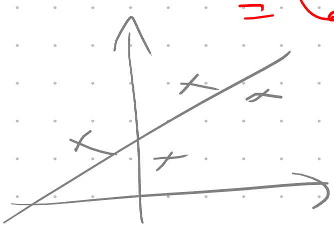
$\varepsilon \sim \cancel{\mathcal{N}}(\text{Pois}(\lambda))$



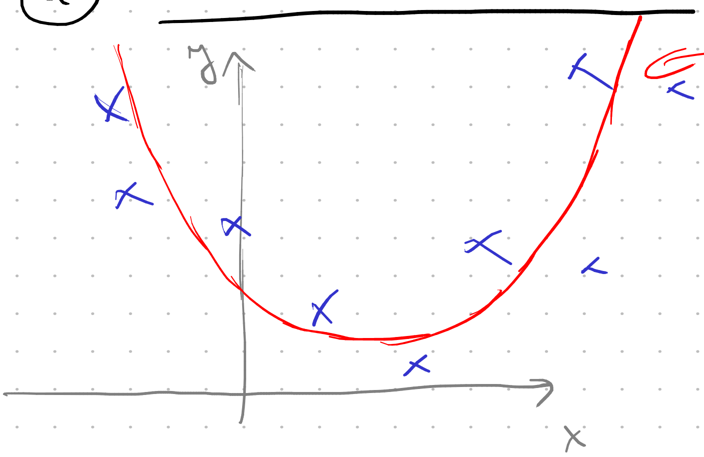
$$\log p(D|\bar{w}) = \text{Const} - \frac{1}{2}\sum_{n=1}^{N}\frac{1}{\sigma_n^2}(y_n - \bar{w}^T\bar{x}_n)^2 =$$

$$= \text{Const} - \frac{1}{2}\Big(\frac{1}{\sigma_1^2}\sum_{n: x_n < a}(y_n - \bar{w}^T\bar{x}_n)^2 + \frac{1}{\sigma_2^2}\sum_{n: x_n > a}(y_n - \bar{w}^T\bar{x}_n)^2\Big)$$

$\to \min$



4 Feature extraction in LR



$$\hat{y} = w_0 + w_1 x + w_2 x^2 \qquad \xrightarrow[\bar{w}]{} \min$$

$$L(\bar{w}) = \sum_n \big(y_n - (w_0 + w_1 x_n + w_2 x_n^2)\big)^2$$

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

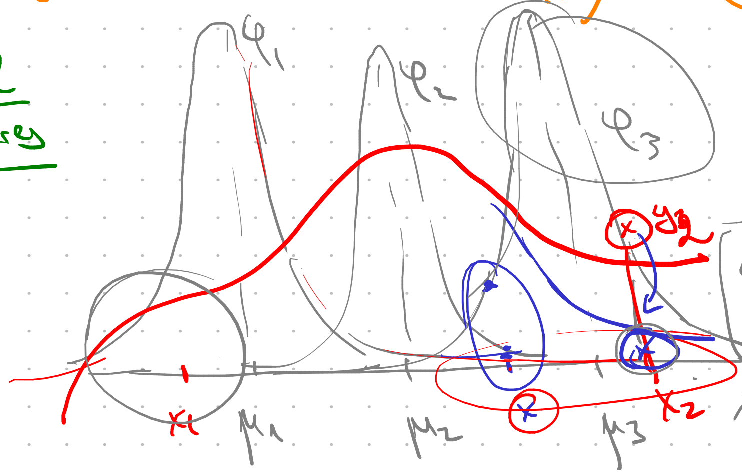
feature extraction

$\bar{x}$ — NN — $\bar{\varphi}(\bar{x})$ — Lin. regr. — $y = \bar{w}^T \bar{\varphi}(\bar{x})$

polynomial regression

$$x \xrightarrow{\bar{\varphi}} \begin{pmatrix} 1 \\ x \\ \dots \\ x^k \end{pmatrix}^T \begin{pmatrix} w_0 \\ w_1 \\ \dots \\ w_k \end{pmatrix}$$

local features

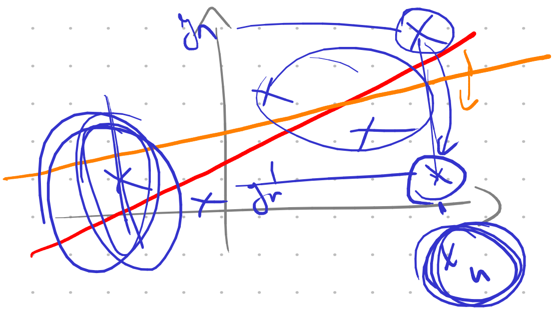


$$x \mapsto \begin{pmatrix} 1 \\ \varphi_1(x) \\ \varphi_2(x) \\ \varphi_3(x) \end{pmatrix}$$

$$\varphi_j(x) = c \cdot e^{-c' \cdot (x-\mu_j)^2}$$

radial basis functions

RBF



$$\hat{y} = w_0 + w_1 \varphi_1(x) + w_2 \varphi_2(x) + w_3 \varphi_3(x)$$