① Logistic regression

$\{(\bar{x}_n, t_n)\}_{n=1}^{N}, \quad t_n = \begin{cases} 1, & C_1 \\ 0, & C_2 \end{cases}$

$p(C_1|\bar{x}) = \sigma(\bar{w}^T\bar{x})$

$p(C_2|\bar{x}) = 1 - \sigma(\bar{w}^T\bar{x}), \quad \sigma(a) = \frac{1}{1+e^{-a}}$

$$p(D|\bar{w}) = \prod_{n=1}^{N} p(t_n|\bar{x}_n, \bar{w}) = \prod_{n=1}^{N} \sigma(\bar{w}^T\bar{x}_n)^{t_n} \left(1 - \sigma(\bar{w}^T\bar{x}_n)\right)^{1-t_n} \qquad \bar{w} \to \max$$

$$\log p(D|\bar{w}) = \sum_{n=1}^{N} \left[t_n \log \sigma(\bar{w}^T\bar{x}_n) + (1-t_n)\log(1-\sigma(\bar{w}^T\bar{x}_n)) \right]$$

$t_n - t_n\sigma_n - (1-t_n)\sigma_n = t_n - \cancel{t_n\sigma_n} - \sigma_n + \cancel{t_n\sigma_n}$

$$\nabla_{\bar{w}} \log p(D|\bar{w}) = \sum_{n=1}^{N} \Big[t_n(1-\underbrace{\sigma(\bar{w}^T\bar{x}_n)}_{\sigma_n})\bar{x}_n - (1-t_n)\cdot\sigma(\bar{w}^T\bar{x}_n)\bar{x}_n \Big] =$$

$\sigma'(a) = \frac{1 \cdot e^{-a}}{(1+e^{-a})^2} = \sigma(a)\cdot(1-\sigma(a))$

$(\log\sigma)' = \frac{\sigma'}{\sigma} = 1-\sigma$

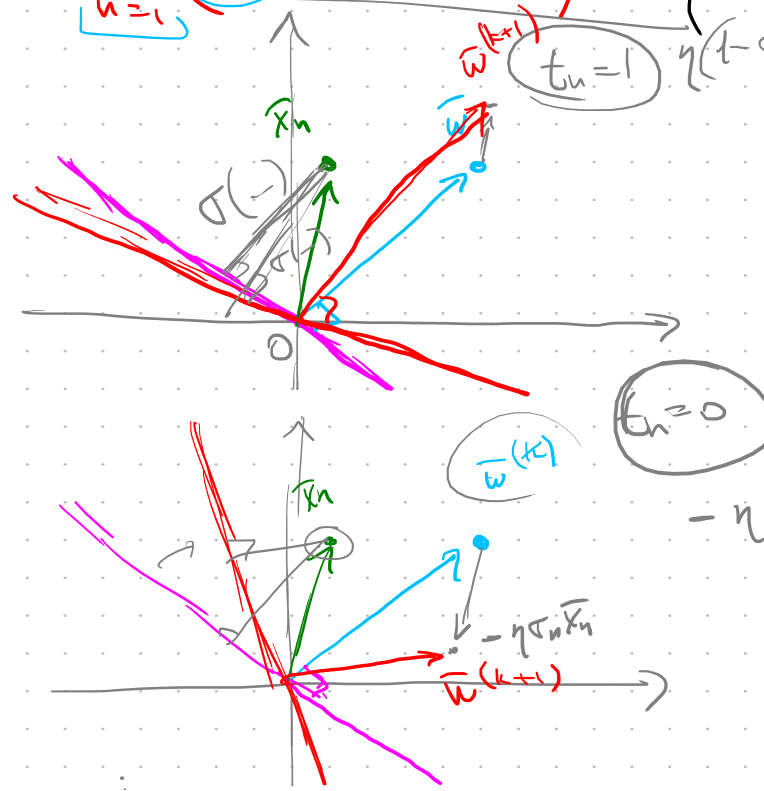
$\log(1-\sigma)' = \frac{-\sigma'}{1-\sigma} = -\sigma$

$$= \sum_{n=1}^{N} \left(t_n - \sigma(\bar{w}^T\bar{x}_n)\right)\bar{x}_n \;\; - \frac{1}{\sigma_0^2}\bar{w} \;\; + \log p(\bar{w}) \;\; - \frac{1}{2\sigma_0^2}\bar{w}^T\bar{w}$$

$\eta(1-\sigma_n)\cdot\bar{x}_n$

$p(C_1|\bar{x}) = \sigma(\bar{w}^T\bar{x}) = 1/2$

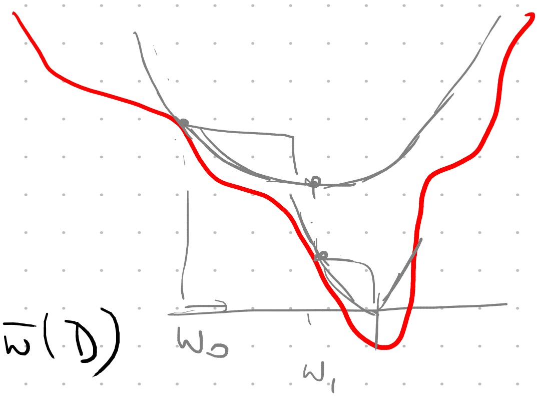
$\bar{w}^T\bar{x} = 0$



SGD – stochastic grad. descent

$\bar{w} := \bar{w} + \eta \cdot \nabla_{\bar{w}} \log p(x_n|\bar{w})$

$-\eta\cdot\sigma\cdot\bar{x}_n$



$p(D|\bar{w}) \times p(\bar{w}) \quad \propto \quad p(\bar{w}|D)$

$\text{Const} + \log p(D|\bar{w}) + \log p(\bar{w}) = \log p(\bar{w}|D)$

max ↑

$$f(\bar{w}) \approx f(\bar{w}_0) + \nabla_{\bar{w}} f(\bar{w}_0)^T(\bar{w} - \bar{w}_0) + \frac{1}{2}(\bar{w}-\bar{w}_0)^T H (\bar{w}-\bar{w}_0)$$

$$H = \left(\frac{\partial^2 f}{\partial w_i \partial w_j} \right)$$

hessian

$$\nabla_{\bar{w}}(\ ) = \nabla_{\bar{w}} f(\bar{w}_0) + H(\bar{w}_* - \bar{w}_0) = 0$$

$$\boxed{\bar{w}_* = w_0 - H^{-1} \cdot \nabla_{\bar{w}} f(\bar{w}_0)}$$

$$\nabla_{\bar{w}} \log p(D|\bar{w}) = \sum_n (t_n - \sigma(\bar{w}^T \bar{x}_n)) \cdot \bar{x}_n$$

$$\frac{\partial^2 \log p(D|\bar{w})}{\partial w_i \partial w_j} = \frac{\partial}{\partial w_i} \left(\sum_n (t_n - \sigma(\bar{w}^T \bar{x}_n)) \cdot x_{nj} \right) =$$

$$= - \sum_{n=1}^{N} \sigma(\bar{w}^T \bar{x}_n)(1 - \sigma(\bar{w}^T \bar{x}_n)) \cdot x_{ni} x_{nj}$$

j

$$\sum_n \sigma_n(1-\sigma_n) x_{ni} x_{nj} \quad i$$

$$X^T X$$

$$H = -X^T \cdot \underbrace{\begin{pmatrix} \sigma_1(1-\sigma_1) & & 0 \\ & \sigma_2(1-\sigma_2) & \\ 0 & & \ddots \\ & & \sigma_N(1-\sigma_N) \end{pmatrix}}_{R} \cdot X$$

$$\bar{w}^{(k+1)} = \bar{w}^{(k)} + (X^T R X)^{-1} \cdot X^T (\bar{t} - \bar{\sigma}(\bar{w}^{(k)T} \bar{x}))$$

LR

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

IRLS - iterative reweighted least squares

Pred. distr.

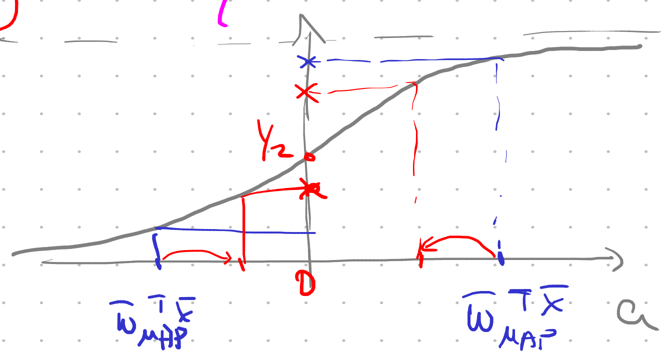
$$p(C_1|\bar{x}, D) = \int p(C_1|\bar{w}, \bar{x})\, p(\bar{w}|D)\, d\bar{w} = $$

$$= \int \sigma(\bar{w}^T \bar{x}) \cdot p(\bar{w}|D)\, d\bar{w} \approx$$

$$\approx \sigma\left(\frac{\boxed{\bar{w}_{MAP}^T \bar{x}}}{\sqrt{1 + \frac{\pi}{8} \bar{x}^T \Sigma_N \bar{x}}} \right)$$

$\to 0$

Bishop



(2) Multiclass logistic regression

$W = (\bar{w}_k)$ $\quad D = \{(\bar{x}_n, \bar{t}_n)\}_{n=1}^{N}$

one-hot

$$p(C_k \mid \bar{w}_k, \bar{x}_n) = \frac{e^{\bar{w}_k^T \bar{x}_n}}{\sum_s e^{\bar{w}_s^T \bar{x}_n}} = y_{nk}$$

$\bar{t}_n = (0 \ldots 0, 1, 0 \ldots 0)$

$t_{nk} = \begin{cases} 1, & \bar{x}_n \in C_k \\ 0, & - \end{cases} \in \mathbb{R}^{|C|}$

$$p(D \mid W) = \prod_{n=1}^{N} \prod_{k=1}^{K} p(C_k \mid \bar{w}_k, \bar{x}_n)^{t_{nk}}$$

$$\log p(D \mid W) = \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \cdot \log y_{nk}$$

1, true

" 0, false

$(y_k - y_j)$ $\quad y_k = \frac{e^{a_k}}{\sum_s e^{a_s}}$, $\quad (a_1 \ldots a_K)$

$$\frac{\partial y_k}{\partial a_k} = \frac{e^{a_k} \cdot \sum_s - e^{a_k} \cdot e^{a_k}}{(\sum_s)^2} = \frac{e^{a_k}}{(\sum_s)^2} - \left(\frac{e^{a_k}}{\sum_s}\right)^2 = y_k(1 - y_k)$$

$$\frac{\partial y_k}{\partial a_j} = y_k \cdot ([k=j] - y_j)$$

$$\frac{\partial y_k}{\partial a_j} = -\frac{e^{a_k} \cdot e^{a_j}}{(\sum_s a^s)^2} = -y_k \cdot y_j$$

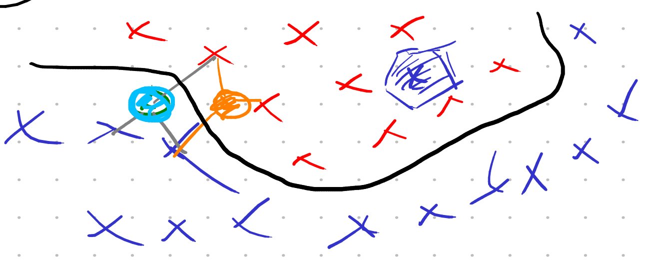
$$\nabla_{\bar{w}_j} \log p(D \mid W) = \sum_n \sum_k t_{nk} \cdot \nabla_{\bar{w}_j} \log y_{nk}$$

$$= \sum_n \sum_k t_{nk} \cdot \frac{\cancel{y_{nk}} \cdot ([k=j] - y_{nj})}{\cancel{y_{nk}}} \cdot \bar{x}_n$$

$$= \sum_n \left(\sum_k t_{nk}([k=j] - y_{nj})\right) \bar{x}_n = \sum_n (t_{nj} - y_{nj}) \bar{x}_n$$

$$\sum_k t_{nk}([k=j] - y_{nj}) = t_{nj} - \Big(\underbrace{\sum_k t_{nk}}_{=1}\Big) y_{nj}$$

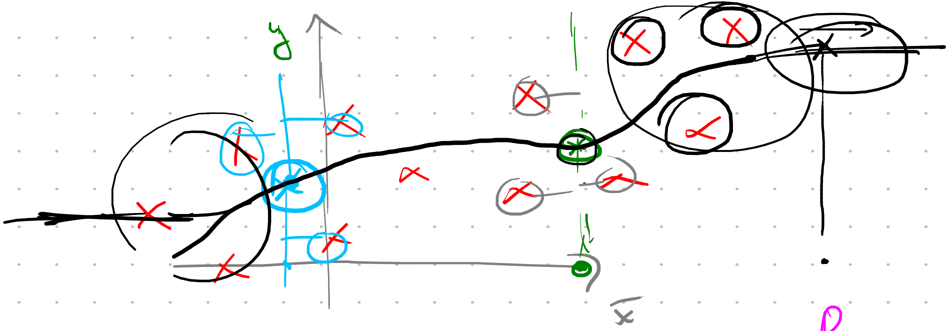
(3) NN & the curse of dimensionality

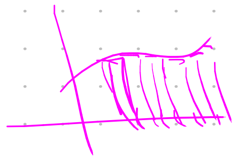


k-NN - k nearest neighbors

k=1

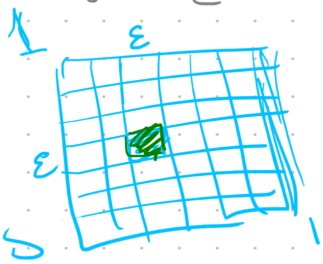
non-parametric method





1) Numerical integration $\int_a^b f(\bar{x})\,d\bar{x} \approx \sum \dots$

$d=1$ $\quad {}_0[\dots]_1 \quad \varepsilon \quad 1/\varepsilon$



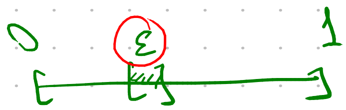
$d=2$ $\quad 1/\varepsilon^2$

$d=d$ $\quad 1/\varepsilon^d$

Sampling $\quad p(\bar{x}|D) = E\left[p(\bar{x}|\bar{\theta})\right]_{p(\bar{\theta}|D)}$

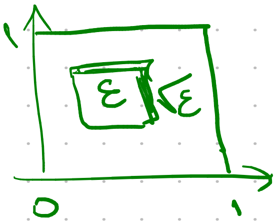
$\approx \frac{1}{R}\sum \dots \quad \bar{\theta}_r \sim p(\bar{\theta}|D)$

2) $d=1$

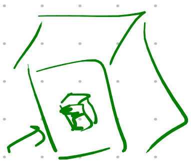


$d:$ $\quad \sqrt[d]{\varepsilon} = \varepsilon^{1/d}$

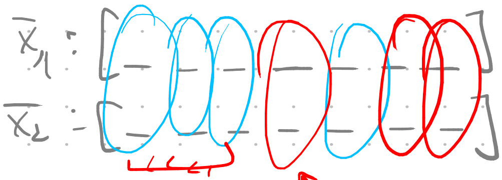
$d=2$



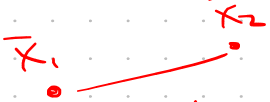
$d=3$ $\quad \sqrt[3]{\varepsilon}$



$N = 1/\varepsilon$



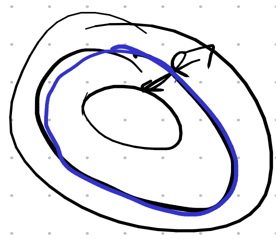
important



$d(\bar{x}_1,\bar{x}_2)^2 = \sum_{i=1}^{d} (x_{1i} - x_{2i})^2$

Kernel trick

SVM



$\bar{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \xrightarrow{\bar{\varphi}} \begin{pmatrix} x_1 \\ \vdots \\ x_d \\ x_1^2 \\ \vdots \\ x_d^2 \end{pmatrix}$ $\quad \frac{d(d+1)}{2} \sim$ degree k $\quad O(d^k)$

$k(\bar{x}_n, \bar{x}_m)$

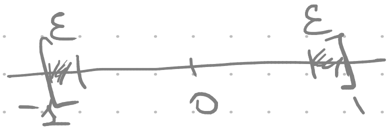
$\dots \bar{x}_n^T \bar{x}_m$

kernel trick: $\bar{\varphi}: \bar{x} \mapsto \bar{\varphi}(\bar{x}) \in \mathbb{R}^D$

$\bar{\varphi}(\bar{x}_1)^T \bar{\varphi}(\bar{x}_2) = k(\bar{x}_1, \bar{x}_2) = e^{-c\cdot(\bar{x}_1 - \bar{x}_2)^2}$

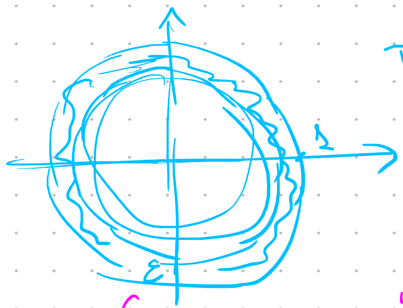
3) Orange peel effect

$d=1$



$$\varepsilon = \frac{2\varepsilon}{2}$$

$d=2$



πr^2

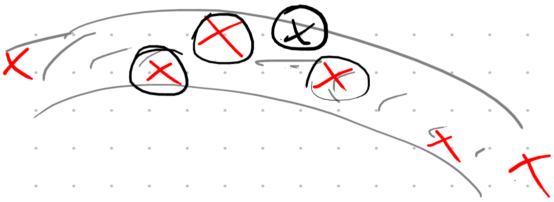
$$\frac{\pi \cdot 1^2 - \pi(1-\varepsilon)^2}{\pi \cdot 1^2} = 1-(1-\varepsilon)^2 = 2\varepsilon - \varepsilon^2$$

$d=3$

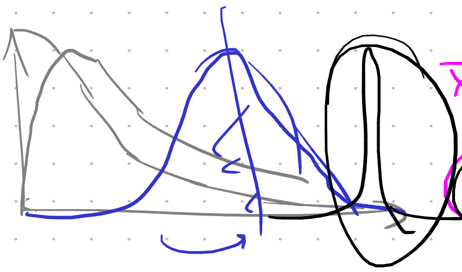
$V = \frac{4}{3}\pi r^3$

$$\frac{\frac{4}{3}\pi \cdot 1^3 - \frac{4}{3}\pi(1-\varepsilon)^3}{4/3\,\pi \cdot 1^3} = 1-(1-\varepsilon)^3$$

d: $\quad 1-(1-\varepsilon)^d \xrightarrow[d\to\infty]{} 1$



$\bar{x}$: [_ _ - - _ _]



$\bar{x} \sim \mathcal{N}(\bar{x} \mid \bar{0}, I)$

$r^2 = \Sigma x_i^2$

$\sim \mathcal{N}(x_i \mid 0, 1)$