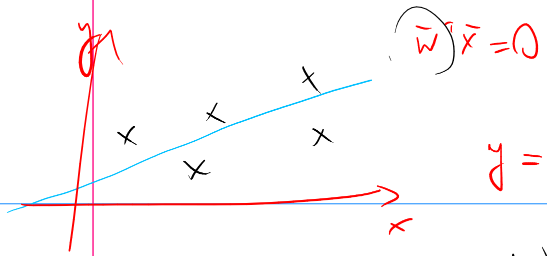



$$y = \bar{w}^T \bar{x} + \varepsilon$$

$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$x \mapsto \begin{pmatrix} 1 \\ x \end{pmatrix} \; ; \; x \mapsto \begin{pmatrix} 1 \\ x \\ \dots \\ x^d \end{pmatrix}$$

$$y = w_0 + w_1 x + \dots + w_d x^d + \varepsilon$$

$$\bar{w} \sim \mathcal{N}(0, \sigma'^2 I)$$

$$\bar{w}_{ML} = \arg\max p(D|\bar{w})$$

$$\bar{w}_{MAP} = \arg\max \left[ p(\bar{w}) \cdot p(D|\bar{w}) \right]$$

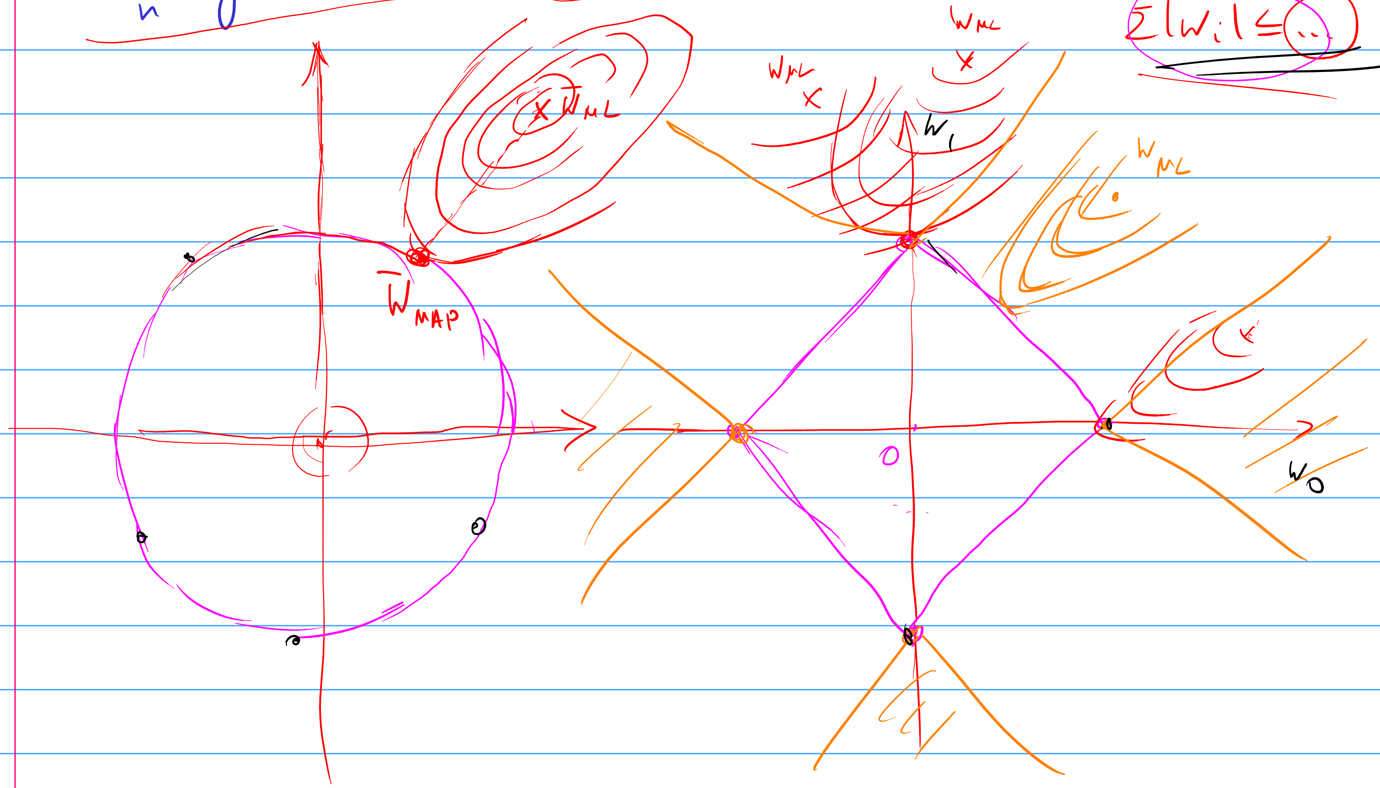
---

$$\bar{w}_{MAP} = \arg\min \left[ \sum_n (y_n - \bar{w}^T \bar{x}_n)^2 + \alpha \sum w_i^2 \right] \Leftrightarrow \min \sum_n \dots, \; \sum w_i^2 \le (\dots)$$

$$\alpha = \frac{\sigma^2}{\sigma'^2} \qquad \sum w_i^2 = \|\bar{w}\|^2$$

---

$$\sum_n (y_n - \bar{w}^T \bar{x}_n)^2 + \alpha \sum |w_i| \to \min \Leftrightarrow \min \sum_n (\dots)^2, \; \sum |w_i| \le (\dots)$$




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

---

$$p(\bar{w}) = \mathcal{N}(\bar{w} | \bar{\mu}_0, \Sigma_0)$$

$$p(\bar{w}|D) \propto p(\bar{w})\, p(D|\bar{w}) = \mathcal{N}(\bar{w} | \bar{\mu}_0, \Sigma_0) \cdot \prod_n \mathcal{N}(y_n | \bar{w}^T \bar{x}_n, \sigma^2)$$

$$\ln p(\bar{w}|D) = \text{Const} - \frac{d}{2} \ln(2\pi) - \frac{1}{2} \ln \det \Sigma_0 - \frac{1}{2} (\bar{w} - \bar{\mu}_0)^T \Sigma_0^{-1} (\bar{w} - \bar{\mu}_0) - \frac{N}{2} \ln(2\pi\sigma^2) - \sum_{n=1}^{N} \frac{1}{2\sigma^2} (y_n - \bar{w}^T \bar{x}_n)^2$$

$$-\frac{1}{2} \bar{w}^T \Sigma_0^{-1} \bar{w}$$

$$-\frac{1}{2\sigma^2} (\bar{y} - X\bar{w})^T (\bar{y} - X\bar{w})$$

$$\ln p(\bar w|D) = \text{Const} - \frac{1}{2}(\bar w - \bar\mu_N)^T \Sigma_N^{-1} (\bar w - \bar\mu_N) \quad \overset{\shortparallel}{=} -\frac{1}{2\sigma^2}\bar y^T \bar y - \frac{1}{2\sigma^2}\bar w^T X^T X \bar w + \frac{1}{\sigma^2}\bar w^T (X^T y)$$

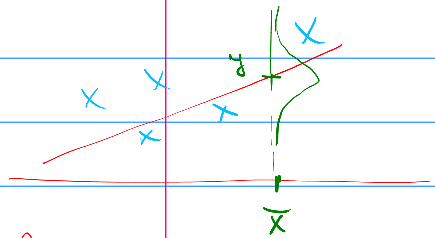
$$-\frac{1}{2}\bar w^T \Sigma_N^{-1} \bar w$$

$$-\frac{1}{2}\bar w^T \Sigma_N^{-1}\bar w = -\frac{1}{2}\bar w^T \Sigma_0^{-1}\bar w - \frac{1}{2\sigma^2}\bar w^T X^T X \bar w$$

$$\boxed{\Sigma_N^{-1} = \Sigma_0^{-1} + \frac{1}{\sigma^2}X^T X}$$

Упр: $\mu_N = ?$

$$p(\bar w) = \mathcal{N}(\bar w|\mu_0, \Sigma_0) \times p(D|\bar w) \propto p(\bar w|D) = \mathcal{N}(\bar w|\mu_N, \Sigma_N)$$




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

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

$$\mathcal{N}(y|\bar w^T \bar x, \sigma^2) \cdot \mathcal{N}(\bar w|\mu_N, \Sigma_N)$$

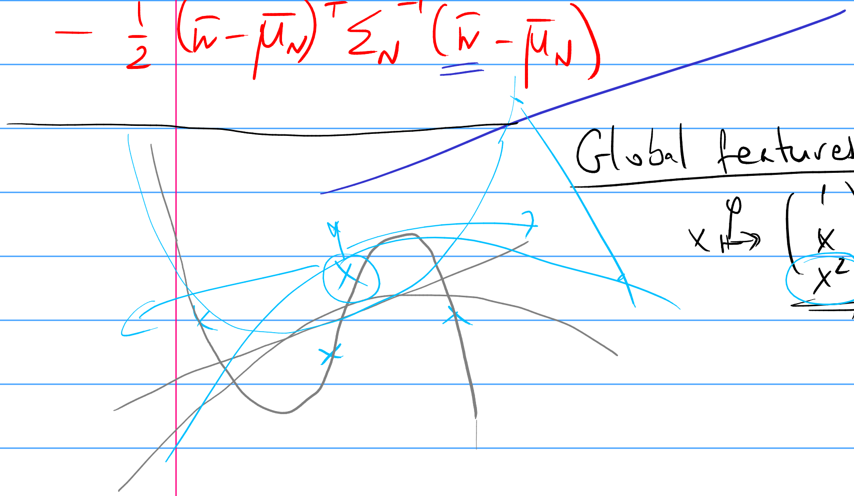
$$\ln(..) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y - \bar w^T \bar x)^2 - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln\det\Sigma_N - \frac{1}{2}(\bar w - \bar\mu_N)^T \Sigma_N^{-1}(\bar w - \bar\mu_N)$$

$$\int \mathcal{N}(\bar w|\_)\, d\bar w = 1$$

$$\mathcal{N}(\bar w|\ldots,\ldots)\cdot(\ldots) = (\ldots) = f(y)$$

$$-\frac{1}{2}(\bar w - \bar\mu')\Sigma'^{-1}(\bar w - \bar\mu') - f(y)$$

Global features

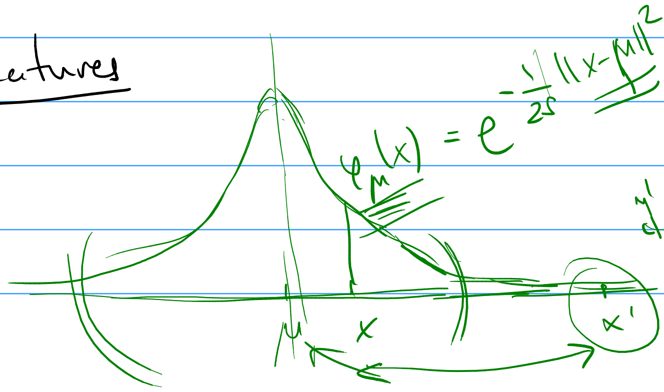


$$x \overset{\varphi}{\mapsto} \begin{pmatrix}1\\ x\\ x^2\end{pmatrix}$$

$$\begin{pmatrix}x\\ y\end{pmatrix} \mapsto \begin{pmatrix}1\\ x\\ y\\ x^2\\ y^2\\ xy\end{pmatrix}$$

$$\begin{pmatrix}x_1\\ \ldots\\ x_d\end{pmatrix} \mapsto \begin{pmatrix}\vdots\\ d^2\\ \vdots\end{pmatrix}$$

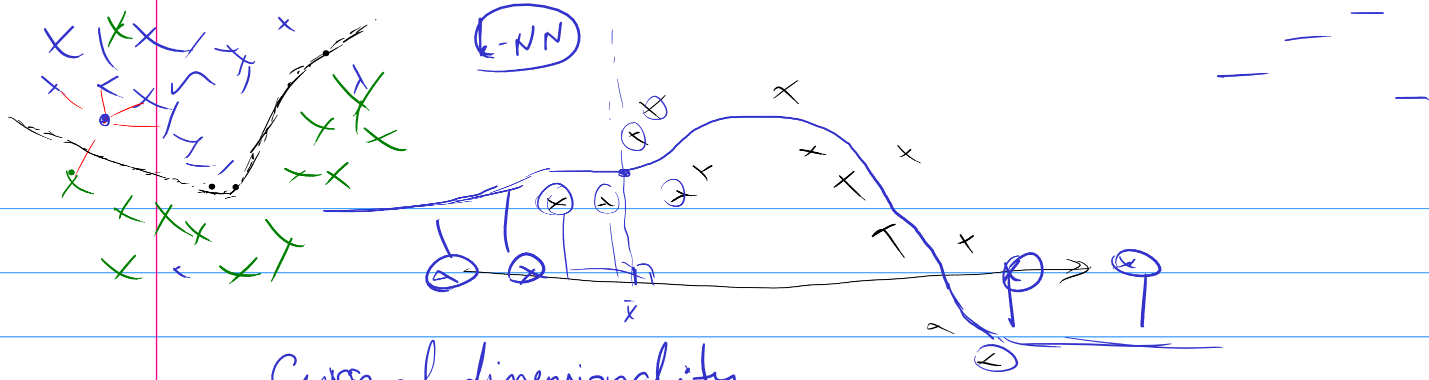
Local features



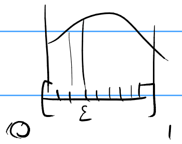

$$\varphi_\mu(x) = e^{-\frac{1}{2s}\|x-\mu\|^2}$$

$$y' = \bar w^T \varphi(\bar x') + \varepsilon$$

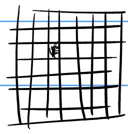
$$\begin{pmatrix}\vdots\\ \varphi_\mu(\bar x)\\ \vdots\end{pmatrix}$$



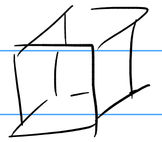

# Curse of dimensionality




$1/\varepsilon$

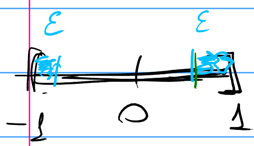


$1/\varepsilon^2$



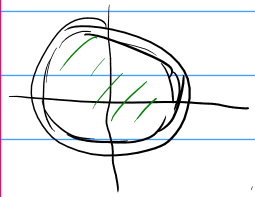
$1/\varepsilon^d$

d = 1

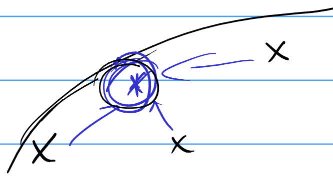



$\frac{2\varepsilon}{2} = \varepsilon$

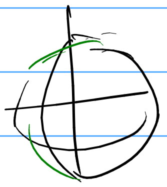
d = 2



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



d = 3



$$\frac{4/3\,\pi - 4/3\,\pi(1-\varepsilon)^3}{4/3\,\pi} = 3\varepsilon - 3\varepsilon^2 + \varepsilon^3$$

d = 1000

d

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

$\bar{x}_1 = ( \ldots x_{1k} \ldots )$

$\bar{x}_2 = ( \ldots x_{2k} \ldots )$

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

$r^2 = \sum x_i^2$

$\bar{x} = ( \ldots x_i \ldots x_j \ldots )$