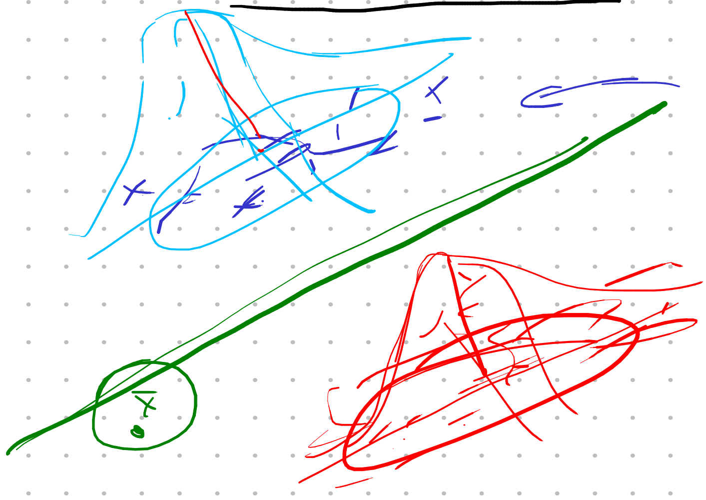
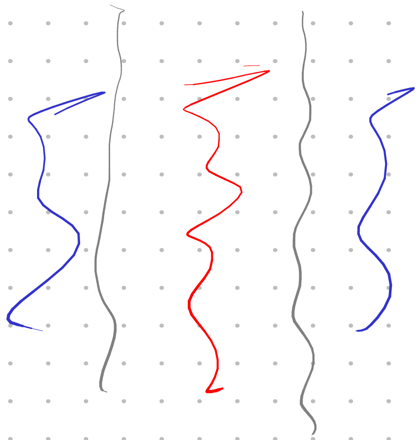
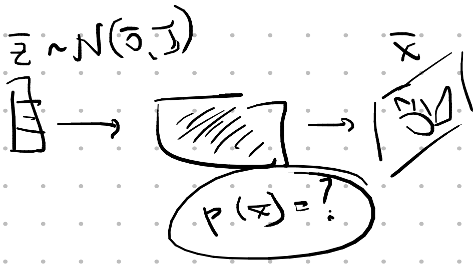
Generative models



$p(\bar{x}|C_1)$

$p(\bar{x}|C_2)$





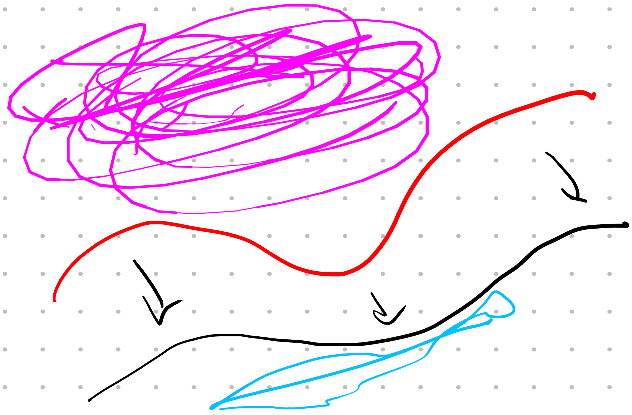
Optimal Bayesian classifier

$$p(C_1|\bar{x}) = \frac{p(\bar{x}|C_1)\,p(C_1)}{p(\bar{x})} = \frac{p(C_1)\,p(\bar{x}|C_1)}{\underbrace{p(C_1)\,p(\bar{x}|C_1)}_{p(\bar{x},C_1)} + \underbrace{p(C_2)\,p(\bar{x}|C_2)}_{p(\bar{x},C_2)}}$$

$$p(C_1|\bar{x}) = p(C_2|\bar{x})$$

$$p(C_1)\,p(\bar{x}|C_1) = p(C_2)\,p(\bar{x}|C_2)$$

$$\log\frac{p(\bar{x}|C_1)}{p(\bar{x}|C_2)} + \log\frac{p(C_1)}{p(C_2)} = 0$$



$$p(\bar{x}|C_1) = \mathcal{N}(\bar{x}|\bar{\mu}_1, \Sigma_1)$$
$$p(\bar{x}|C_2) = \mathcal{N}(\bar{x}|\bar{\mu}_2, \Sigma_2)$$

$$\log p(C_1) - \cancel{\frac{d}{2}\log 2\pi} - \frac{1}{2}\log\det\Sigma_1 - \frac{1}{2}\overbrace{(\bar{x}-\bar{\mu}_1)^T\Sigma_1^{-1}(\bar{x}-\bar{\mu}_1)}^{\bar{\mu}_1^T\Sigma_1^{-1}\bar{\mu}_1 + \bar{x}^T\Sigma_1^{-1}\bar{x} - 2\bar{x}^T\Sigma_1^{-1}\bar{\mu}_1}$$

$$= \log p(C_2) - \cancel{\frac{d}{2}\log 2\pi} - \frac{1}{2}\log\det\Sigma_2 - \frac{1}{2}(\bar{x}-\bar{\mu}_2)^T\Sigma_2^{-1}(\bar{x}-\bar{\mu}_2)$$

$$-\frac{1}{2}\bar{x}^T\left(\Sigma_1^{-1}-\Sigma_2^{-1}\right)\bar{x}+\bar{x}^T\left(\Sigma_1^{-1}\bar{\mu}_1-\Sigma_2^{-1}\bar{\mu}_2\right)-$$

$$-\frac{1}{2}\bar{\mu}_1\Sigma_1^{-1}\bar{\mu}_1+\frac{1}{2}\bar{\mu}_2\Sigma_2^{-1}\bar{\mu}_2-\frac{1}{2}\log\frac{\det\Sigma_1}{\det\Sigma_2}+\log\frac{p(C_1)}{p(C_2)}=0$$

$$p(C_1)=\frac{N_1}{N_1+N_2} \qquad p(C_2)=\frac{N_2}{N_1+N_2}$$

$$\hat{\mu}_1=\frac{1}{N_1}\sum_{\bar{x}\in C_1}\bar{x}$$

$$\bar{x}\sim\mathcal{N}(\bar{\mu},\Sigma)$$

$$\hat{\Sigma}_1=\frac{1}{N_1}\sum_{\bar{x}\in C_1}(\bar{x}-\hat{\mu}_1)(\bar{x}-\hat{\mu}_1)^T \qquad \Sigma^{-1}(\bar{x}-\bar{\mu})\sim\mathcal{N}(\bar{0},I)$$

QDA – quadratic discriminant analysis

LDA – linear DA $\qquad \Sigma_1=\Sigma_2=\Sigma$

$$\bar{x}^T\Sigma^{-1}(\bar{\mu}_1-\bar{\mu}_2)-\frac{1}{2}\bar{\mu}_1^T\Sigma^{-1}\bar{\mu}_1+\frac{1}{2}\bar{\mu}_2^T\Sigma^{-1}\bar{\mu}_2+\log\frac{p(C_1)}{p(C_2)}=0$$

$$\hat{\mu}_1,\ \hat{\mu}_2,\ \underbrace{p(C_1)}_{\text{"}\pi},\ \underbrace{p(C_2)}_{\text{"}1-\pi}$$

$$\boxed{\hat{\Sigma}=\frac{N_1}{N_1+N_2}\hat{\Sigma}_1+\frac{N_2}{N_1+N_2}\hat{\Sigma}_2}$$

$$p(\bar{x}\mid\bar{\mu}_1,\bar{\mu}_2,\Sigma,\pi)=\pi\cdot\mathcal{N}(\bar{x}\mid\bar{\mu}_1,\Sigma)+(1-\pi)\cdot\mathcal{N}(\bar{x}\mid\bar{\mu}_2,\Sigma)$$

$$p(D\mid\pi,\bar{\mu}_1,\bar{\mu}_2,\Sigma)=\prod_{\bar{x}\in C_1}\left[\pi\cdot\mathcal{N}(\bar{x}\mid\bar{\mu}_1,\Sigma_1)\right]\cdot\prod_{\bar{x}\in C_2}\left[(1-\pi)\cdot\mathcal{N}(\bar{x}\mid\bar{\mu}_2,\Sigma_2)\right]\xrightarrow[\pi,\mu_1,\mu_2,\Sigma]{}\max$$

$$E_{p(\bar{x})}[f(\bar{x})]\approx\frac{1}{N}\sum_{\bar{x}_n\sim p(\bar{x})}f(\bar{x}_n)$$

$$\log p(D|\ldots) = \sum_{\bar x \in C_1} \left(\log \pi + \left[Const - \frac{1}{2} \log \det \Sigma_1 - \frac{1}{2} (\bar x - \bar \mu_1)^T \Sigma_1^{-1} (\bar x - \bar \mu_1) \right] \right) + \sum_{\bar x \in C_2} \left(\log(1-\pi) + \left[- \ldots \mu_2 \ \Sigma_2 \right] \right)$$

$\pi \nearrow \max$

$$= \left[\sum_{C_1} \log \pi + \sum_{C_2} \log(1-\pi) \right] + \ldots$$

$\longrightarrow \max$

$\pi, \mu_1, \mu_2, \Sigma$

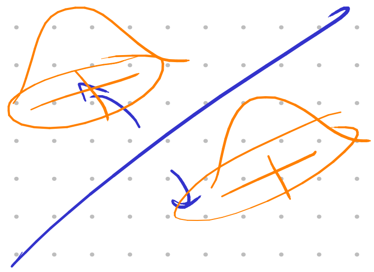
Σ_1, Σ_2

$$\hat\pi = \frac{N_1}{N_1 + N_2} \qquad \hat\mu_1, \ \hat\mu_2, \ \hat\Sigma$$

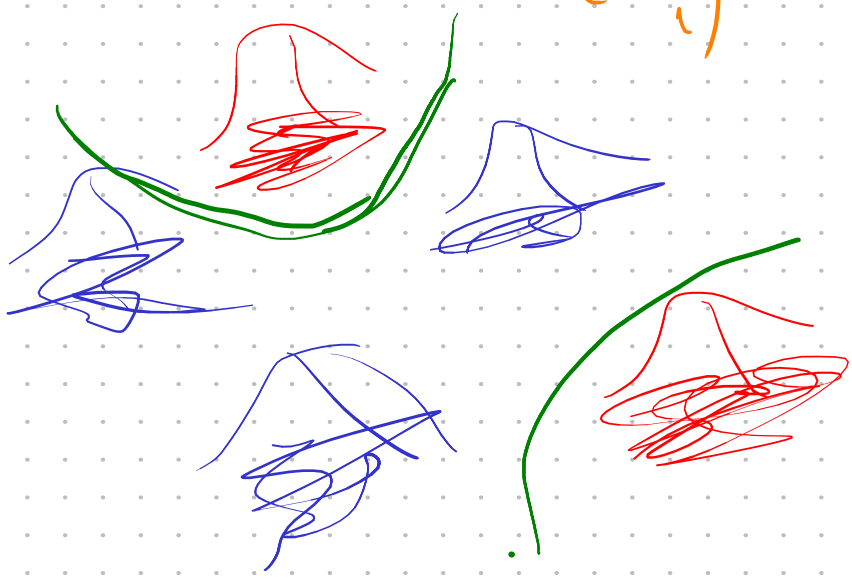
$\mathbb{R}^2$ $\qquad$ $\mathbb{R}^5$

$$\bar x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \longmapsto \begin{pmatrix} x_1 \\ x_2 \\ x_1 x_2 \\ x_1^2 \\ x_2^2 \end{pmatrix}$$

$$\bar w^T \bar\varphi(\bar x) = 0$$



$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \longmapsto \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} 2 \\ 0 \end{pmatrix} \longmapsto \begin{pmatrix} 2 \\ 0 \\ 0 \\ 4 \\ 0 \end{pmatrix}$$



Logistic regression — discriminative model $\cancel{p(\bar x|C_1)}$

$p(C_1|\bar x)$

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

$p(C_1|\bar x)$ $\qquad$ $\cancel{\bar w^T \bar x}$

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

$$t_n = \begin{cases} 1, & \bar x_n \in C_1 \\ 0, & \bar x_n \in C_2 \end{cases}$$

$$p(C_1|\bar{x}) = \frac{p(C_1)\,p(\bar{x}|C_1)}{p(C_1)p(\bar{x}|C_1) + p(C_2)p(\bar{x}|C_2)} =$$

$$= \frac{1}{1 + \frac{p(C_2)p(\bar{x}|C_2)}{p(C_1)p(\bar{x}|C_1)}}$$

$$\frac{p(C_2|\bar{x})}{p(C_1|\bar{x})}$$ – odds (шансы)

$$\cancel{\bar{w}^T\bar{x}}$$

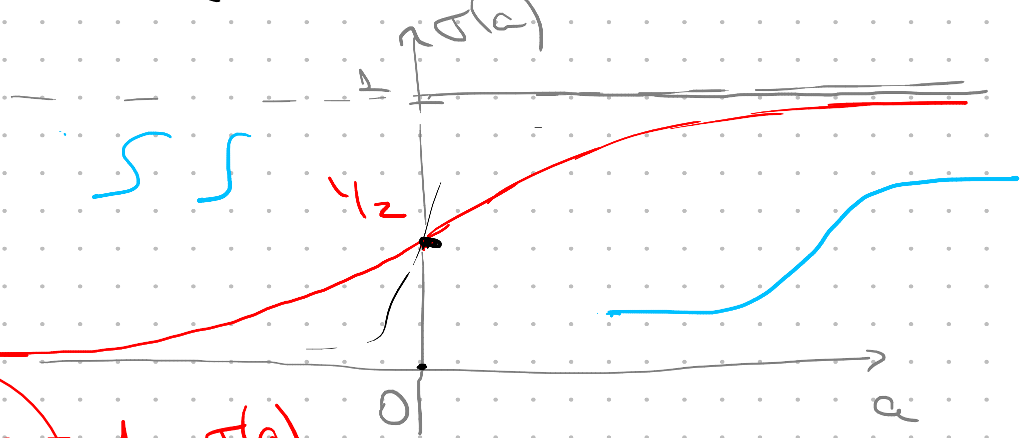
$$= \frac{1}{1 + e^{-\boxed{\log \frac{p(C_1)p(\bar{x}|C_1)}{p(C_2)p(\bar{x}|C_2)}}}} \qquad \approx \bar{w}^T\bar{x} + \cancel{w_0} \qquad = \frac{1}{1+e^{-a}}$$

$$\underbrace{}_{=a}$$

log-odds.

logistic sigmoid

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



tanh

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

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

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

$$p(C_1|\bar{x}) = \frac{1}{2}$$

$$\sigma(\bar{w}^T\bar{x})$$

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

$p(\bar{w})$

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

$$= \prod_n \begin{cases} \sigma(\bar{w}^T\bar{x}_n), & \text{если } t_n = 1 \\ 1-\sigma(\bar{w}^T\bar{x}_n), & \text{если } t_n = 0 \end{cases}$$

$$= \prod_{n=1}^{N} \cancel{\left[t_n \sigma(\bar w^T \bar x_n) + (1-t_n)\left(1-\sigma(\bar w^T \bar x_n)\right) \right]}$$

$$= \prod_{n=1}^{N} \sigma(\bar w^T \bar x_n)^{t_n} \left(1-\sigma(\bar w^T \bar x_n)\right)^{1-t_n} \qquad \boxed{\xrightarrow[\bar w]{} \max}$$

$+\log p(\bar w)$

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

$\sigma(\bar w^T \bar x_n)(1-\sigma(\bar w^T \bar x_n))$

$$\nabla_{\bar w} \log p(D|\bar w) = \sum_{n=1}^{N} \left[t_n \cdot \frac{\sigma'(\bar w^T \bar x_n)}{\sigma(\bar w^T \bar x_n)} \bar x_n - (1-t_n) \frac{\sigma'(\bar w^T \bar x_n)}{1-\sigma(\bar w^T \bar x_n)} \bar x_n \right]$$

$$t_n - \cancel{t_n \sigma_n} - \sigma_n + \cancel{t_n \sigma_n}$$

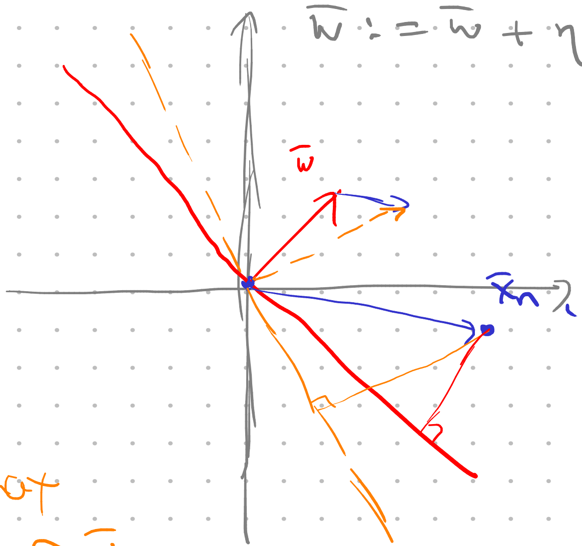
$$= \sum_n \left[t_n\left(1-\sigma(\bar w^T \bar x_n)\right) - (1-t_n)\sigma(\bar w^T \bar x_n) \right] \bar x_n$$

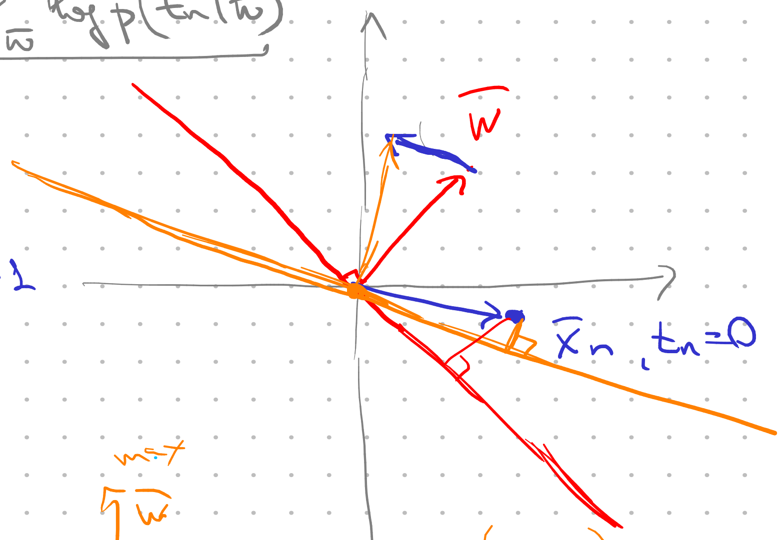
$\bar w \to \max$

$$\nabla_{\bar w} \log p(D|\bar w) = \sum_{n=1}^{N} \left(t_n - \sigma(\bar w^T \bar x_n)\right) \bar x_n$$

stochastic gradient descent

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





max $\uparrow \bar w$

max $\uparrow \bar w$

$-2\bar w^T(H\bar w_0)$

$$f(\bar w) \approx f(\bar w_0) + \nabla_{\bar w} f\Big|_{\bar w_0}^T (\bar w - \bar w_0) + \frac{1}{2}(\bar w - \bar w_0)^T H (\bar w - \bar w_0) = f_1$$

$$H = \left(\frac{\partial^2 f}{\partial w_i \partial w_j}\Big|_{\bar w_0} \right)_{ij}$$

гессиан hessian

$$\nabla_{\bar w} f_1 = \nabla_{\bar w} f\Big|_{\bar w_0} + H\bar w_* - H\bar w_0 = 0$$

$$H \bar w_* = H\bar w_0 - \nabla_{\bar w} f\Big|_{\bar w_0}$$

$$\boxed{\bar{w}_* = w_0 - H^{-1} \cdot \nabla_{\bar{w}} f \Big|_{\bar{w}_0}}$$ — Newton's method

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

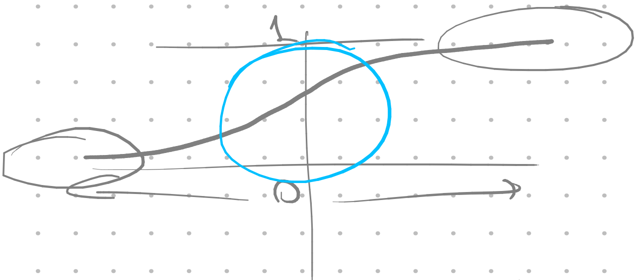
$$\nabla_{\bar{w}} L = (X^T X)\bar{w} - X^T\bar{y} \qquad \bar{w}_* = w_0 - (X^T X)^{-1}(X^T X \bar{w}_0 - X^T \bar{y})$$

$$H = X^T X \qquad\qquad = (X^T X)^{-1} X^T \bar{y}$$

$$\frac{\partial^2 \log p}{\partial w_i \partial w_j} = \frac{\partial}{\partial w_i}\left(\sum_{n=1}^{N}\left(t_n - \sigma(\bar{w}^T \bar{x}_n)\right) x_{nj}\right) =$$

$$= \sum_{n=1}^{N}\left(-\sigma'(\bar{w}^T\bar{x}_n) \cdot x_{ni} \cdot x_{nj}\right) = -\sum_{n=1}^{N} \sigma_n(1-\sigma_n) x_{ni} x_{nj}$$

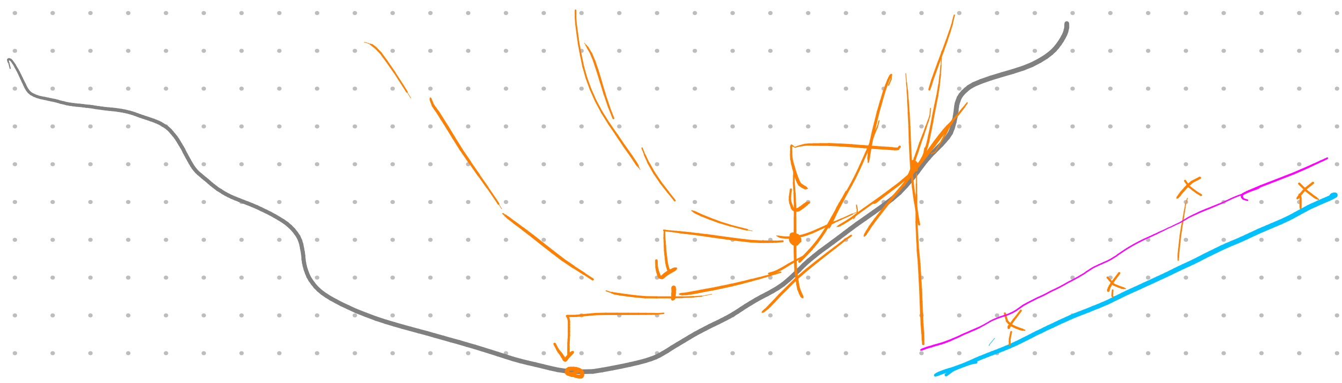
$$H = -\sum_n \sigma_n(1-\sigma_n)\bar{x}_n\bar{x}_n^T = -X^T R X$$



$$R = \begin{pmatrix} \sigma_1(1-\sigma_1) & & 0 \\ & \sigma_2(1-\sigma_2) & \\ 0 & & \ddots \\ & & & \sigma_n(1-\sigma_n) \end{pmatrix}$$

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

IRLS – iterative reweighted least squares

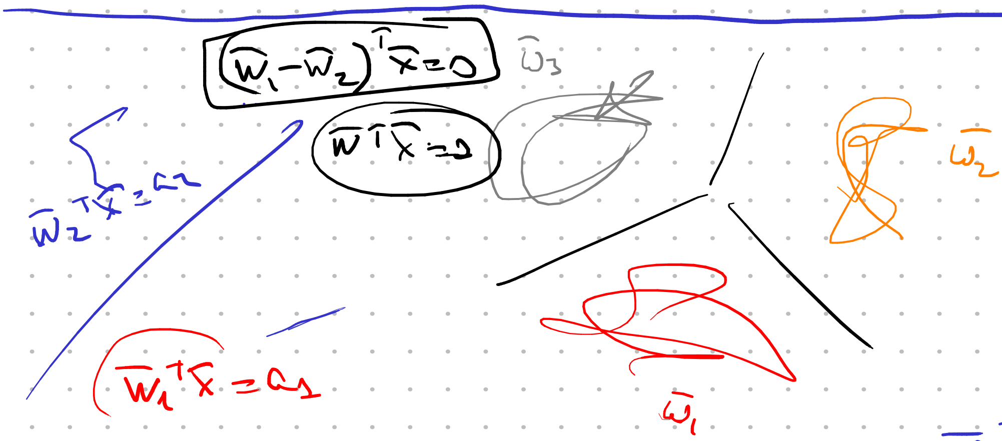


$$p(\bar{w}|D) \propto p(\bar{w})\, p(D|\bar{w})$$

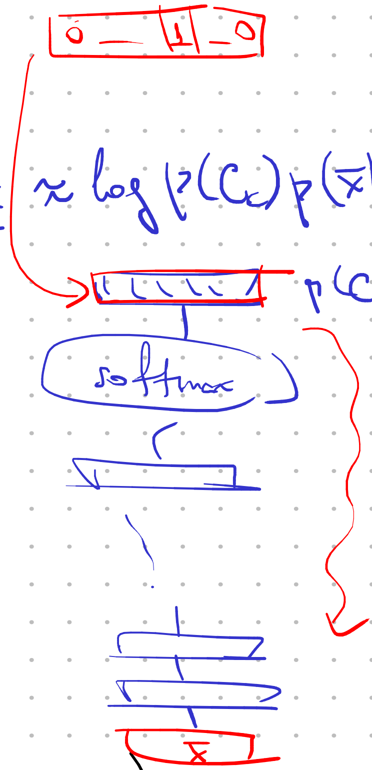
$$\log p(\bar{w}|D) = \text{const} + \sum_n \left[t_n \ldots + (1-t_n) \ldots \right] + \underbrace{\log p(\bar{w})}_{-\frac{1}{2\sigma_0^2}\bar{w}^T\bar{w}}$$

$$\nabla \log p(\bar{w}|D) = \nabla \log p(D|\bar{w}) - \alpha \cdot \bar{w}$$

$$H' = H - \alpha \cdot I$$



$$\bar{w}_k^T \bar{x} \approx \log p(C_k)\, p(\bar{x}|C_k)$$



$$p(C_k|\bar{x}) = \frac{p(C_k)\, p(\bar{x}|C_k)}{\sum_{s=1}^{K} p(C_s)\, p(\bar{x}|C_s)}$$

$$p(C_k|\bar{x}) = \frac{e^{\bar{w}_k^T \bar{x}}}{\sum_s e^{\bar{w}_s^T \bar{x}}}$$

$$\text{softmax}(a_1, \ldots, a_k, \ldots a_K) = \left(\ldots, \frac{e^{a_k}}{\sum_s e^{a_s}}, \ldots \right)$$

$$\bar{t}_n = (0 \ldots 1 \ldots 0) \qquad t_{nk} = 1 \iff \bar{x}_n \in C_k$$

$$p(D|W) = \prod_n p(\bar{t}_n | W, \bar{x}_n) = \prod_{n=1}^{N} \prod_{k=1}^{K} \underbrace{p(C_k|\bar{w}_k, \bar{x}_n)}_{= y_{nk}}{}^{t_{nk}}$$

$$\log p(D|W) = \sum_n \sum_k t_{nk} \log y_{nk} = \text{softmax}_k(\bar{w}_1^T \bar{x}_n, \ldots, \bar{w}_K^T \bar{x}_n) = e^{\bar{w}_k^T \bar{x}_n} / \sum e^{\bar{w}_i^T \bar{x}_n}$$

$$y_k = \frac{e^{a_k}}{\sum_i e^{a_i}} \qquad \frac{\partial y_k}{\partial a_j} = \begin{cases} \dfrac{e^{a_k}\sum_i e^{a_i} - e^{a_k}\cdot e^{a_k}}{(\sum_i a_i)^2} = y_k(1 - y_k), & k=j \\[2ex] -\dfrac{e^{a_k}\cdot e^{a_j}}{(\sum_i e^{a_i})^2} = -y_k \cdot y_j, & k \neq j \end{cases} =$$

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

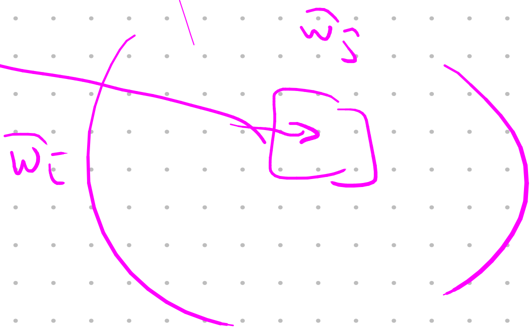
$$\nabla_{\bar{w}_j} \log p(D|W) = \sum_{n=1}^{N} \sum_k t_{nk} \frac{y_{nk}([k=j] - y_{nj})}{y_{nk}} \bar{x}_n$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk}([k=j] - y_{nj})\bar{x}_n =$$

$$= \sum_n \Big[t_{nj}\bar{x}_n - \underbrace{\Big(\sum_k t_{nk}\Big)}_{1} y_{nj}\bar{x}_n \Big] = \sum_n (t_{nj} - y_{nj})\bar{x}_n$$

$$\boxed{\nabla_{\bar{w}_i} \nabla_{\bar{w}_j} \log p(D|W)} = \begin{cases} -\sum_n y_{nj}(1 - y_{nj})\bar{x}_n, & i=j \\ -\sum_n y_{ni} y_{nj} \bar{x}_n, & i \neq j \end{cases}$$

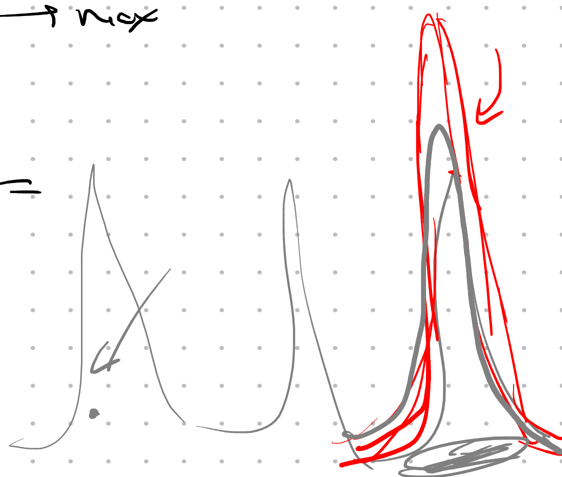
$$W = (\bar{w}_1, \bar{w}_2, \ldots, \bar{w}_K)$$



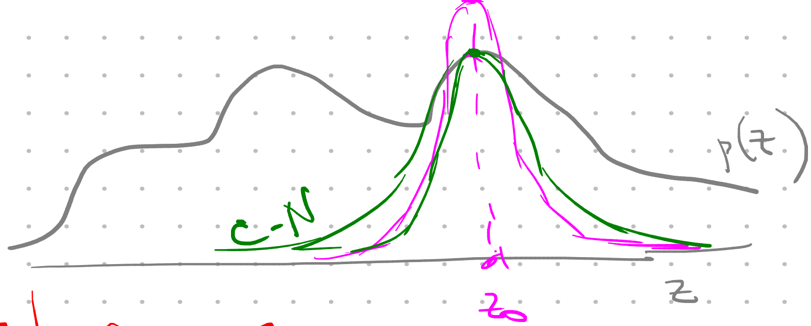
$\log p(D|\bar{w}) \to \max \qquad \log p(\bar{w}|D) \to \max$

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

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



Laplace approximation



$$\log p(z) \approx \log p(z_0) +$$

$$+ \cancel{\left.\frac{\partial \log p}{\partial z}\right|_{z_0}(z-z_0)} + \frac{1}{2}\left.\frac{\partial^2 \log p}{\partial z^2}\right|_{z_0}(z-z_0)^2 =$$

$$= \log p(z_0) - \frac{1}{2}A\cdot(z-z_0)^2, \text{ ye } A = -\left.\frac{\partial^2 \log p}{\partial z^2}\right|_{z_0}$$

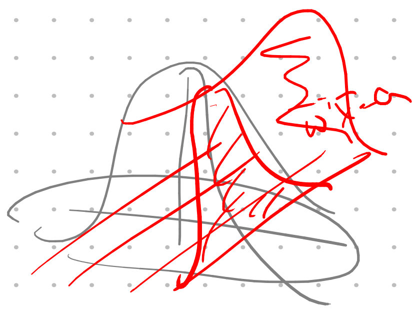
$$p(z) \approx p(z_0)\, e^{-\frac{1}{2}\cdot A(z-z_0)^2}$$

$$\log p(\bar{z}) \approx \log p(\bar{z}_0) - \frac{1}{2}(\bar{z}-\bar{z}_0)^T A(\bar{z}-\bar{z}_0), \text{ ye}$$

$$A = -\nabla\nabla \log p$$

$$p(C_1|D,\bar{x}) = \int \sigma(\bar{w}^T\bar{x})\cdot p(\bar{w}|D)\, d\bar{w} \overset{\bar{w}_{MAP}}{\approx}$$

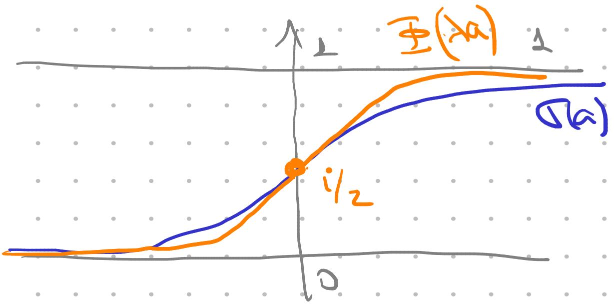
$$\approx \int \sigma(\underbrace{\bar{w}^T\bar{x}}_{a})\cdot \mathcal{N}(\bar{w}|\bar{w}_{MAP}, \Sigma_N)\, d\bar{w} =$$



$$\sigma(\bar{w}^T\bar{x}) = \int \underline{\delta(\bar{w}^T\bar{x}-a)}\cdot \overline{\sigma(a)}\, da$$

$$= \int_{-\infty}^{\infty} \sigma(a)\cdot\left(\int \delta(\bar{w}^T\bar{x}-a)\cdot \underline{\mathcal{N}(\bar{w}|\bar{w}_{MAP},\Sigma_N)}\, d\bar{w}\right) da =$$

$$= \int \underline{\sigma(a)}\cdot \mathcal{N}(a|\overbrace{\bar{w}_{MAP}^T\bar{x}}^{\mu_a}, \overbrace{\bar{x}^T\Sigma_N\bar{x}}^{\sigma_a^2})\, da \approx$$



$\int_{-\infty}^{b} p(x)\,dx$

probit regression

probit $\Phi(b) = \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx$

$\Phi'(\lambda a)\big|_0 = \sigma'(a)\big|_0 \qquad \lambda = \sqrt{\pi/8}$

$$\approx \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\lambda a} \mathcal{N}(x\,|\,0,1)\cdot\mathcal{N}(a\,|\,\mu_a,\sigma_a^2)\,dx\right) da = y_{np}$$

$$= \Phi\left(\frac{\mu_a}{\sqrt{\frac{1}{\lambda^2}+\sigma_a^2}}\right) = \Phi\left(\frac{\lambda\mu_a}{\sqrt{1+\lambda^2\sigma_a^2}}\right) \approx$$

$\bar{x}$

$$\approx \sigma\left(\frac{\mu_a}{\underbrace{\sqrt{1+\frac{\pi}{8}\sigma_a^2}}_{\kappa}}\right)$$

$\mu_a = \bar{w}_{MAP}^{T}\bar{x}$

$\sigma_a^2 = \bar{x}^{T}\Sigma_N\bar{x}$