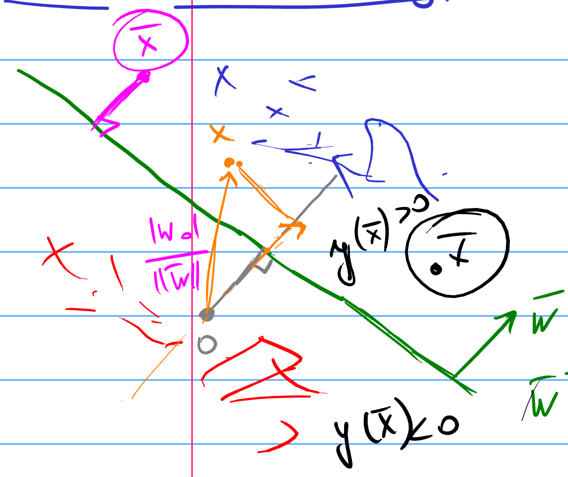
$p(D|\theta) \to \max$ $\qquad$ $\prod p(d|\theta) \to \max$

$p(\theta)\, p(D|\theta) \to \max$ $\qquad$ $p(\theta) \prod p(d|\theta) \to \max$

$$\mathbb{E}_{p(x,y)}[f] = \int p(x) f(x)\, dx$$



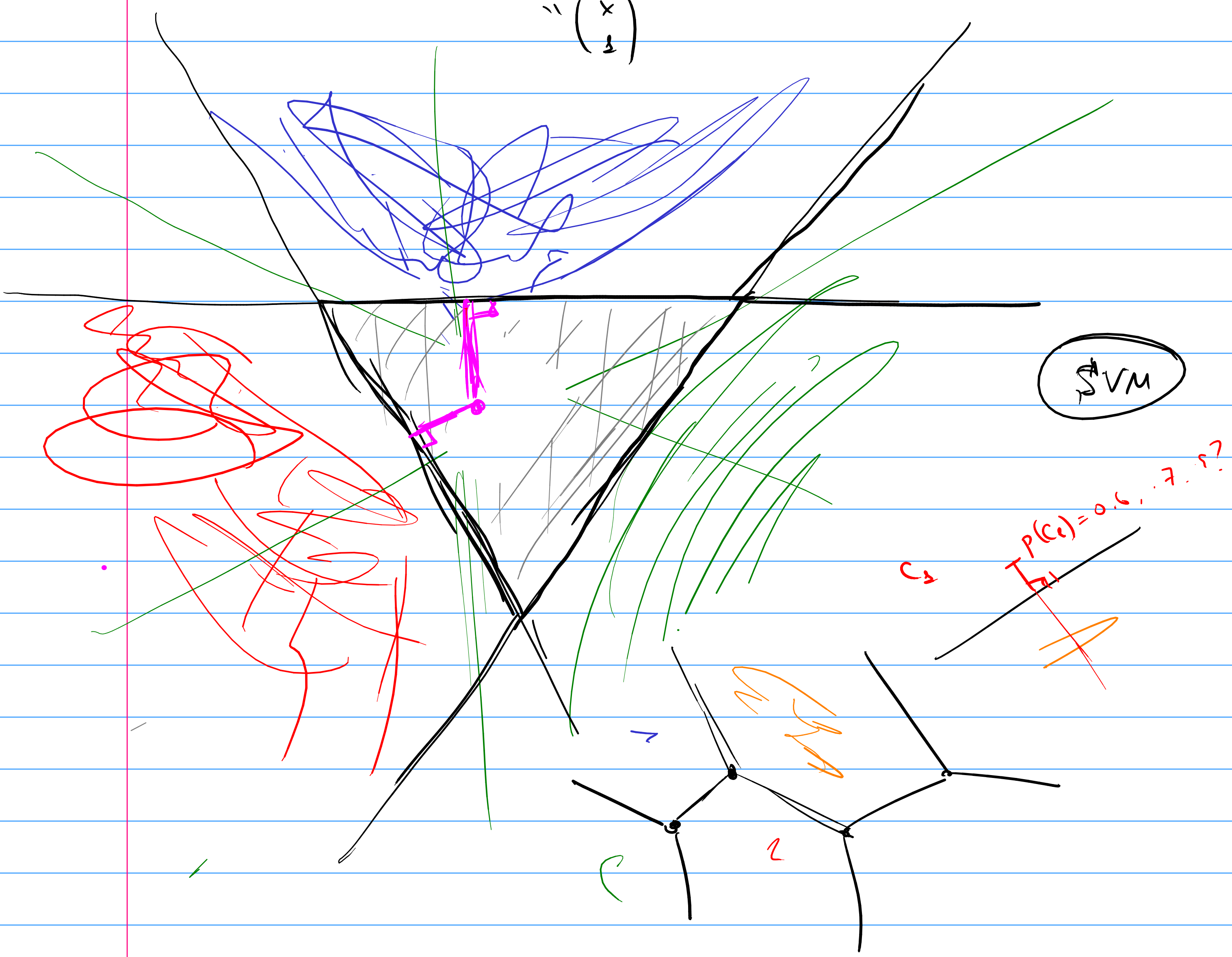
$y(\bar{x}) = \bar{w}^T \bar{x} + w_0$

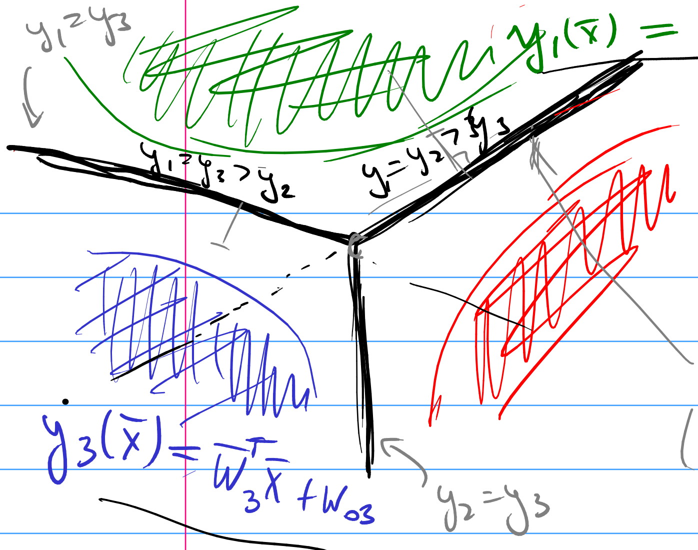
$$d(\bar{x}, \bar{w}^T \bar{x} + w_0) = \frac{|y(\bar{x})|}{\|\bar{w}\|} = \frac{|\bar{w}^T \bar{x} + w_0|}{\|\bar{w}\|}$$

$\bar{w}^T \bar{x} + w_0 = 0$

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

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





$y_1(\bar{x}) = \bar{w}_1 \bar{x} + w_{01}$

$y_2(\bar{x}) = \bar{w}_2 \bar{x} + w_{02}$

$y_3(\bar{x}) = \bar{w}_3^T \bar{x} + w_{03}$

$\operatorname{argmax}\{y_1, y_2, y_3\}$

$y_1 = y_2$

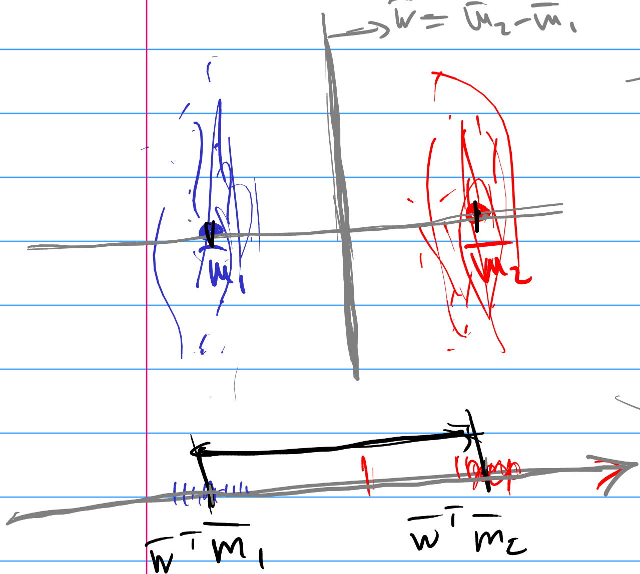
$(\bar{w}_1 - \bar{w}_2)^T \bar{x} + w_{01} - w_{02} = 0$

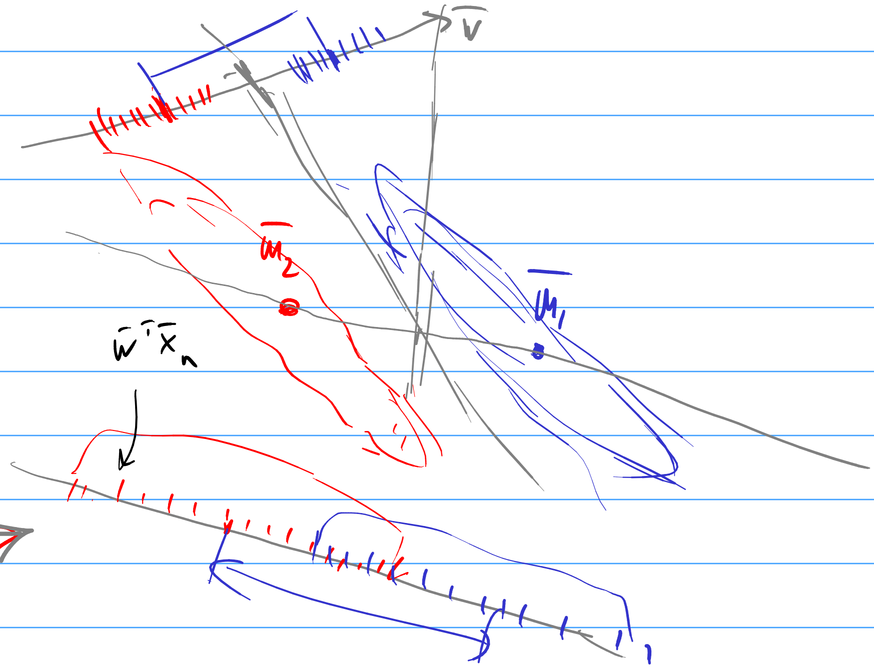
tropical prevarieties

(min, +)

$\max\{y_1, y_2, y_3\}$ дост. не в 1 точке

Fischer linear discriminant





$\bar{w}: \quad \| \bar{w}^T (\bar{m}_1 - \bar{m}_2) \| \to \max$

$S_1 + S_2 \to \min \qquad S_B$ – between-class covariance

$$J(\bar{w}) = \frac{\bar{w}^T (\bar{m}_1 - \bar{m}_2)(\bar{m}_1 - \bar{m}_2)^T \bar{w}}{\sum_{\bar{x} \in C_1} (\bar{w}^T \bar{x} - \bar{w}^T \bar{m}_1)^2 + \sum_{\bar{x} \in C_2} (\bar{w}^T \bar{x} - \bar{w}^T \bar{m}_2)^2} \xrightarrow[\bar{w}]{} \max$$

$$S_W = \sum_{\bar{x} \in C_1} (\bar{x} - \bar{m}_1)(\bar{x} - \bar{m}_1)^T + \sum_{\bar{x} \in C_2} (\bar{x} - \bar{m}_2)(\bar{x} - \bar{m}_2)^T$$

$$J(\bar{w}) = \frac{\bar{w}^T S_B \bar{w}}{\bar{w}^T S_W \bar{w}} \xrightarrow[\bar{w}]{} \max \ ?$$

within-class covariance

$$S_B + S_W = S = \sum_{\bar{x}} (\bar{x}-\bar{m})(\bar{x}-\bar{m})^T \qquad \bar{m} = \mathrm{avg}(\bar{x})$$

$$\nabla_{\bar{w}} J = \frac{2S_B\bar{w}\cdot(\bar{w}^T S_W \bar{w}) - 2S_W\bar{w}(\bar{w}^T S_B \bar{w})}{- - -} = 0$$

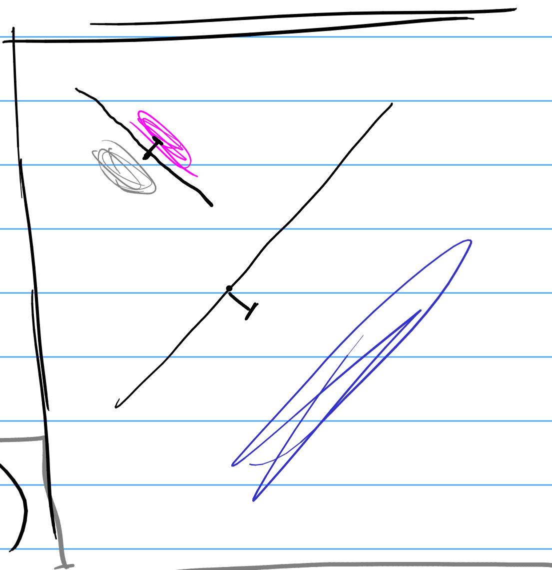
$$\cancel{(\bar{w}^T S_W \bar{w})}\, S_B\bar{w} = \cancel{(\bar{w}^T S_B \bar{w})}\, S_W\bar{w}$$

$$S_B\bar{w} \propto S_W\bar{w}$$

$$(\bar{m}_1-\bar{m}_2)\cancel{(\bar{m}_1-\bar{m}_2)^T\bar{w}} \propto S_W\bar{w}$$

$$S_W\bar{w} \propto \bar{m}_1-\bar{m}_2$$

$$\boxed{\bar{w} \propto S_W^{-1}(\bar{m}_1-\bar{m}_2)}$$

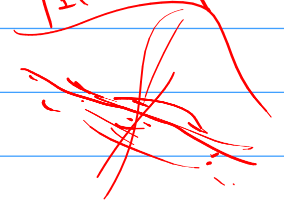


$$\boxed{p(C_1|\bar{x}) = ?}$$

Generative

$p_1(\bar{x}) = p(\bar{x}|C_1)$ $\quad$ $p_2(\bar{x}) = p(\bar{x}|C_2)$

$\bar{x}$?



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

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

optimal Bayesian classifier

$$p(C_k|\bar{x}) = \frac{p(C_k)p(\bar{x}|C_k)}{\sum_\ell p(C_\ell)p(\bar{x}|C_\ell)}$$

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

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

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

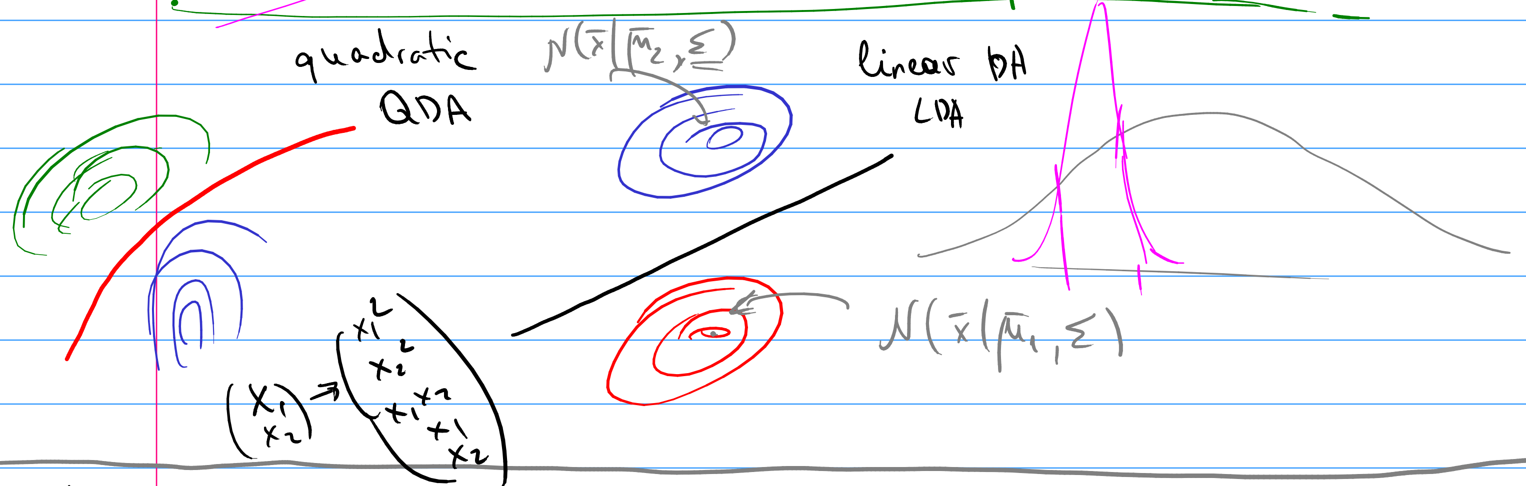
$$\ln p(C_1) + \ln p(\bar{x}|C_1) = \ln p(C_2) + \ln p(\bar{x}|C_2)$$

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

$$\frac{1}{2}\ln|\det\Sigma_1| - \cancel{\frac{d}{2}\ln 2\pi} - \frac{1}{2}(\bar{x}-\bar{\mu}_1)^T\Sigma_1^{-1}(\bar{x}-\bar{\mu}_1) - \frac{1}{2}\ln\det\Sigma_2 + \cancel{\frac{d}{2}\ln 2\pi}$$

$$+\frac{1}{2}(\bar{x}-\bar{\mu}_2)\Sigma_2^{-1}(\bar{x}-\bar{\mu}_2) + \ln\frac{p(C_1)}{p(C_2)} = 0$$

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



Discriminative

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

$$p(C_1|\bar{x}) = \frac{p(C_1)p(\bar{x}|C_1)}{p(C_1)p(\bar{x}|C_1)+p(C_2)p(\bar{x}|C_2)} = \frac{1}{1+\frac{p(C_2)p(\bar{x}|C_2)}{p(C_1)p(\bar{x}|C_1)}} \quad ?\,\bar{w}^T\bar{x}$$

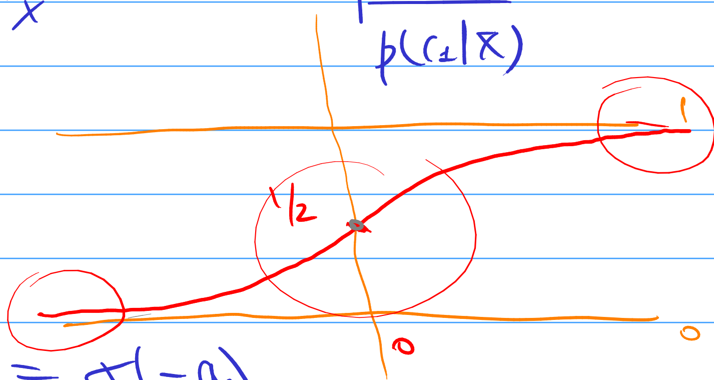
odds

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

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

log-odds $\approx \bar{w}^T\bar{x}$

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



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

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

$$\cancel{t_n p_1 + (1-t_n)p_2}$$

$t_n \in \{0,1\}$ $D=\{(\bar{x}_n, t_n)\}$

$$p(D|\bar{w}) = \prod p(d|\bar{w}) = \prod_n p(t_n|\bar{w},\bar{x}_n) = \prod_n \begin{cases} p(c_1|\_), & t_n=1 \\ p(c_2|\_), & t_n=0 \end{cases}$$

$$= \prod_n p(c_1|\bar{w},\bar{x}_n)^{t_n} \, p(c_2|\bar{w},\bar{x}_n)^{1-t_n} =$$

$$= \prod_n \sigma(\bar{w}^T\bar{x}_n)^{t_n} \left(1-\sigma(\bar{w}^T\bar{x}_n)\right)^{1-t_n} \xrightarrow[\bar{w}]{} \max$$

$$\ln p(D|\bar{w}) = \sum_n t_n \ln \sigma(\bar{w}^T\bar{x}_n) + (1-t_n)\ln\left(1-\sigma(\bar{w}^T\bar{x}_n)\right)$$

$$\nabla_{\bar{w}} \ln \sigma(\bar{w}^T\bar{x}_n) = \frac{\cancel{\sigma_n}(1-\sigma_n)}{\cancel{\sigma_n}}\bar{x}_n = (1-\sigma_n)\bar{x}_n$$

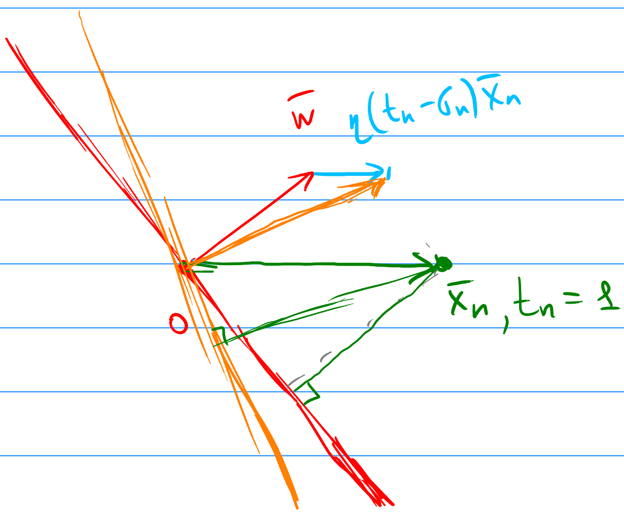
$$\nabla_w \ln(1-\sigma_n) = \frac{-\sigma_n(1-\sigma_n)}{1-\sigma_n}\bar{x}_n = -\sigma_n\bar{x}_n$$

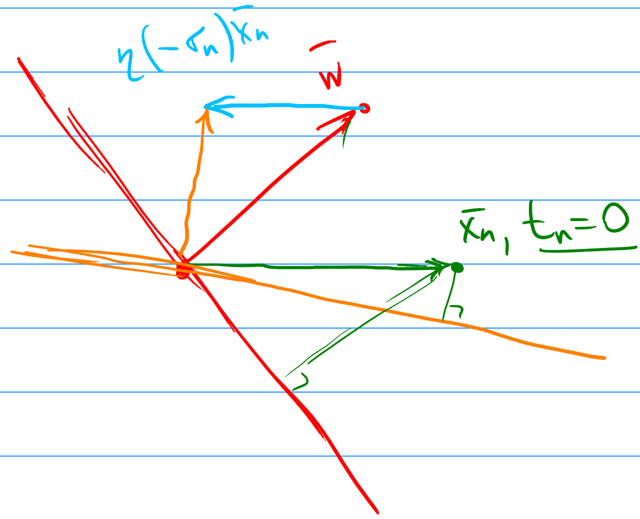
$$\nabla_{\bar{w}} \ln p(D|\bar{w}) = \sum_n \left(t_n(1-\sigma_n)\bar{x}_n - (1-t_n)\sigma_n\bar{x}_n\right) =$$

$$t_n\bar{x}_n - \cancel{t_n\sigma_n\bar{x}_n} - \sigma_n\bar{x}_n + \cancel{t_n\sigma_n}\bar{x}_n$$

$$= \sum_n \boxed{\left(t_n - \sigma(\bar{w}^T\bar{x}_n)\right)\bar{x}_n} \quad + \nabla_{\bar{w}} \ln p(\bar{w})$$

$$\bar{w} := \bar{w} + \eta \cdot \nabla_w \ln p(D|\bar{w})$$





$$\ln p(\bar{w}|D) = \ln p(D|\bar{w}) + \ln p(\bar{w}) + \text{const} \longrightarrow \max$$

$$\sum_n \left[t_n \ln \sigma_n + (1-t_n)\ln(1-\sigma_n)\right] + \ln p(\bar{w}) \longrightarrow \max$$

$$\frac{\partial \ln p(D|\bar w)}{\partial w_i \partial w_j} = \frac{\partial\left(\sum_n \left(t_n - \sigma(\bar w^T \bar x_n)\right) x_{ni}\right)}{\partial w_j}$$

$$\nabla_{\bar w}\left(\sum_n \left(\cancel{t_n} - \sigma(\bar w^T \bar x_n)\right) x_{ni}\right) = -\sum_n \boxed{\sigma_n(1-\sigma_n)}\, x_{ni}\, \bar x_n$$

$$H = \nabla\nabla \ln p(D|\bar w) = \begin{array}{c} \\ i \end{array}\!\!\begin{array}{c} j \\ \left(- \sum_n \sigma_n(1-\sigma_n) x_{ni} x_{nj} \right) \end{array} =$$

$$= X^T \begin{pmatrix} \ddots & & 0 \\ & -\sigma_n(1-\sigma_n) & \\ 0 & & \ddots \end{pmatrix} X = X^T A X$$

"A"

IRLS

iterative reweighted LS

$$\ln p(D|\bar w) = \ldots - \frac{1}{2\sigma^2}(\bar y - X\bar w)^T(\bar y - X\bar w)$$

y max

$$= \ldots - \bar w^T (X^T X)\bar w \ldots$$

$$\nabla_{\bar w} \ln p(D|\bar w) =$$

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

$$H = -X^T X$$

$$p(C_1|\bar x) = \frac{p(C_1)\,p(\bar x|C_1)}{p(C_1)\,p(\bar x|C_1) + \ldots + p(C_K)\,p(\bar x|C_K)} = e^{\ln p(C_1)p(\bar x|C_1)}$$

$\approx \bar w_1^T \bar x$

W

$K \times d$

$$p(C_1|\bar x) = \frac{e^{\bar w_1^T \bar x}}{\sum_k e^{\bar w_k^T \bar x}}$$

$\bar t = (0 \ldots 1 \ldots 0)$

C_k

$$\mathrm{softmax}(a_1 \ldots a_K) = \left(\ldots \frac{e^{a_i}}{\sum_l e^{a_l}} \ldots\right)$$

$$p(D|W) = \prod_n \prod_k p(C_k|\bar x_n)^{t_{nk}} \longrightarrow \max$$

$$\ln p(D|W) = \sum_n \sum_k t_{nk}\left(\bar w_k^T \bar x_n - \ln\left(\sum_i e^{\bar w_i^T \bar x}\right)\right) \longrightarrow \max$$

$\bar w_{ML}$: $\ln p(D|\bar w) \longrightarrow \max$

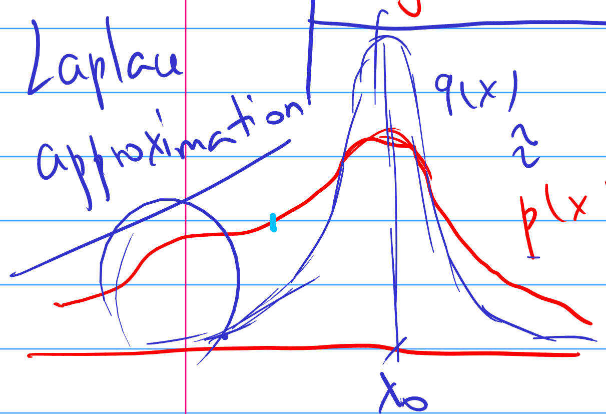
$\bar w_{MAP}$: $\ln p(D|\bar w) + \ln p(\bar w) \longrightarrow \max$

$$\approx q(\bar w) = \mathcal{N}(\bar w | \bar w_{MAP}, \Sigma)$$

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

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

Laplace approximation

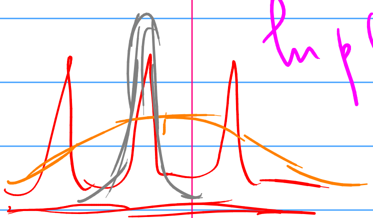


$$q(x) = \dots e^{-\frac{1}{2\sigma^2}(x-x_0)^2}$$

$p(x)$ - -

$KL(p||q) \to \min$

$KL(q||p) \to \min$

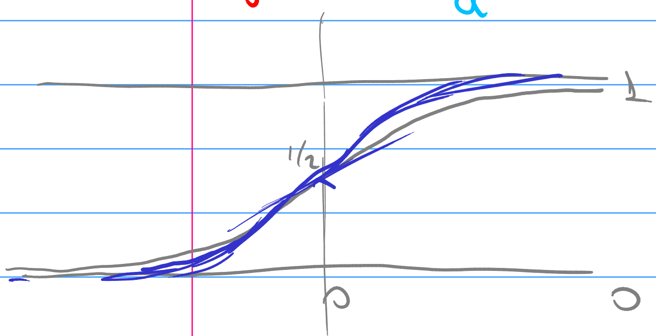


$$\ln p(x) \approx \ln p(x_0) + \cancel{(\ln p(x_0))' \cdot (x-x_0)} + \frac{1}{2}(\ln p(x_0))''(x-x_0)^2$$

$$p(x) \approx p(x_0)\, e^{\frac{1}{2}(\ln p(x_0))''(x-x_0)^2}$$

$$\ln p(\bar x) \approx \ln p(\bar x_0) + \frac{1}{2}(\bar x - \bar x_0)^T \nabla\nabla(\ln p(\bar x_0)) \cdot (\bar x - \bar x_0)$$

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



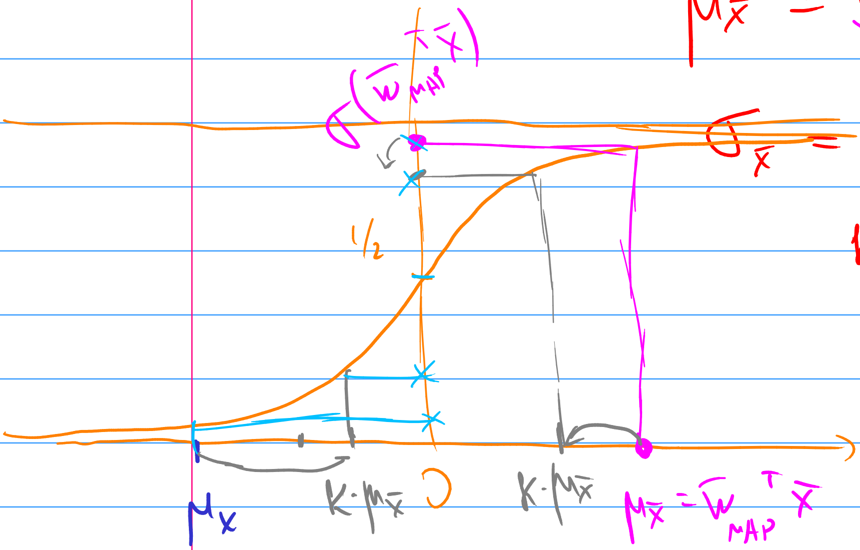
$$F(a) = \int_{-\infty}^{a} p(x)dx \approx \int_{-\infty}^{a} \mathcal{N}(x|\_)dx$$

$$\approx \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{N}(\_)\mathcal{N}(\_)\, da\, dx =$$

$$= \Phi(\_) \approx \sigma(\_)$$

$$p(C_1 \mid \bar{x}, D) \approx \sigma\left(\kappa(\sigma_{\bar{x}}^2)\, \mu_{\bar{x}}\right), \text{ ye}$$

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



$$\sigma_{\bar{x}} = \bar{x}^T \Sigma_N \bar{x}$$

$$\kappa = \frac{1}{\sqrt{1 + \frac{\pi}{8}\sigma_x^2}}$$

$\bar{x}$