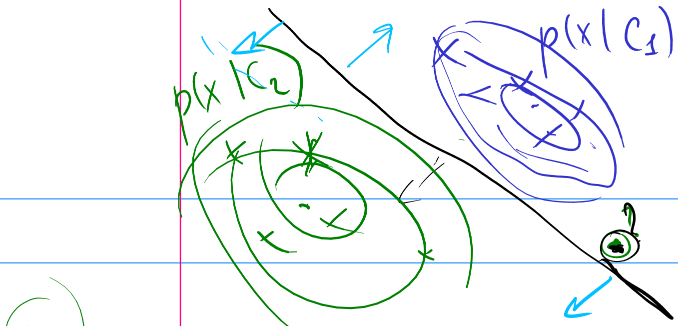


Optimal Bayes classifier

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

$$p(C_1|x)=p(C_2|x)$$

$$p(x|C_1)p(C_1)=p(x|C_2)p(C_2)$$

generative model

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

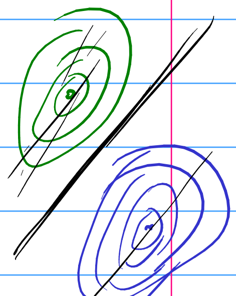
$$\ln p(x|C_1)=\ln p(x|C_2)+\log\frac{p(C_2)}{p(C_1)}$$

$$-\frac{1}{2}\bar{x}^T\Sigma^{-1}\bar{x}$$

1) $\Sigma_1=\Sigma_2=\Sigma$

Linear discr analysis

LDA



$$\cancel{-\frac{1}{2}\ln(2\pi)}\cancel{-\frac{1}{2}\ln\det\Sigma}-\frac{1}{2}(\bar{x}-\bar{\mu}_1)^T\Sigma^{-1}(\bar{x}-\bar{\mu}_1)=$$

$$=\cancel{-\frac{1}{2}\ln 2\pi}\cancel{-\frac{1}{2}\ln\det\Sigma}-\frac{1}{2}(\bar{x}-\bar{\mu}_2)^T\Sigma^{-1}(\bar{x}-\bar{\mu}_2)+\ln\frac{p(C_2)}{p(C_1)}$$

$$\bar{x}^T(\Sigma^{-1}\bar{\mu}_1)-\frac{1}{2}\bar{\mu}_1^T\Sigma^{-1}\bar{\mu}_1=$$

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

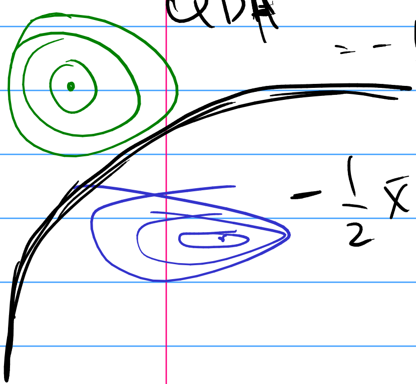
$\ln p(D|\bar{\mu}_1,\bar{\mu}_2,\Sigma)\to\max$

$$\boxed{\Sigma=\frac{N_1}{N_1+N_2}\Sigma_1+\frac{N_2}{N_1+N_2}\Sigma_2}$$

$$\bar{x}^T\Sigma^{-1}(\bar{\mu}_1-\bar{\mu}_2)+C=0$$

2) $\Sigma_1\neq\Sigma_2$

QDA



$$-\frac{1}{2}\ln\det\Sigma_1-\frac{1}{2}(\bar{x}-\bar{\mu}_1)^T\Sigma_1^{-1}(\bar{x}-\bar{\mu}_1)=$$

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

$$-\frac{1}{2}\bar{x}^T\underbrace{(\Sigma_1^{-1}-\Sigma_2^{-1})}\bar{x}+\_\_\_\_=0$$

$$\bar{\mu}_1=\frac{1}{N_1}\sum_{C_1}\bar{x}\qquad\bar{\mu}_2=\frac{1}{N_2}\sum_{C_2}\bar{x}$$

$$\Sigma_1=\frac{1}{N}\sum(\bar{x}-\bar{\mu}_1)(\bar{x}-\bar{\mu}_1)^T$$

Logistic regression - discriminative model

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

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

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

odds

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

logistic sigmoid

log-odds

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

$\bar{w}^T\bar{x} \to \infty \quad p(C_1|\bar{x}) \to 1$

$\bar{w}^T\bar{x} \to -\infty \quad p(C_1|\bar{x}) \to 0$

$\sigma(\bar{w}^T\bar{x}) = \frac{1}{2}$

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

$$p(t=1|\bar{x}) = p(C_1|\bar{x}) = \sigma(\bar{w}^T\bar{x}) = \frac{1}{1+e^{-\bar{w}^T\bar{x}}}$$

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

$t_n \in \{0,1\}$

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

$$\sum_n \ln\left((1-t_n)+(2t_n-1)\sigma(\bar{w}^T\bar{x}_n)\right) \longrightarrow \max$$

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

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

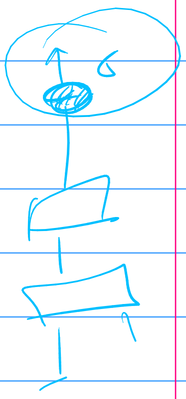
$$(\ln\sigma(a))' = 1-\sigma(a) \qquad (\ln(1-\sigma(a)))' = -\sigma(a)$$

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

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

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

$$= \sum_n \left[t_n - \cancel{t_n\sigma(\bar{w}^T\bar{x}_n)} - \sigma(\bar{w}^T\bar{x}_n) + \cancel{t_n\sigma(\ldots)} \right] \bar{x}_n$$

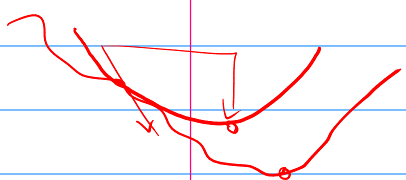


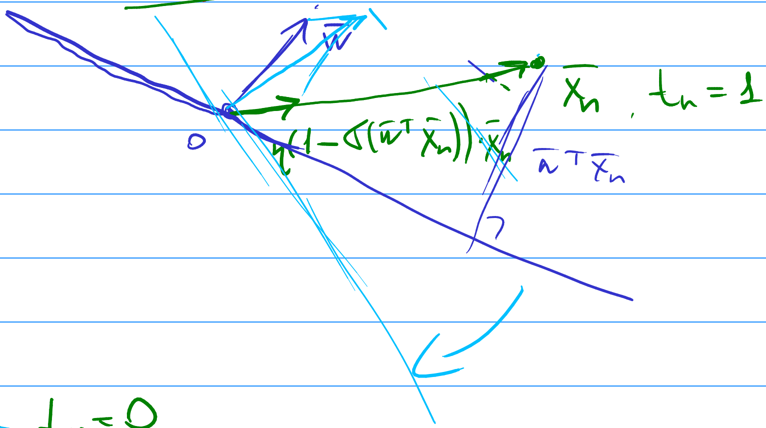
$$\nabla_{\bar{w}} \ln p(D|\bar{w}) = \sum_n \left[t_n - \sigma(\bar{w}^T\bar{x}_n) \right] \bar{x}_n$$

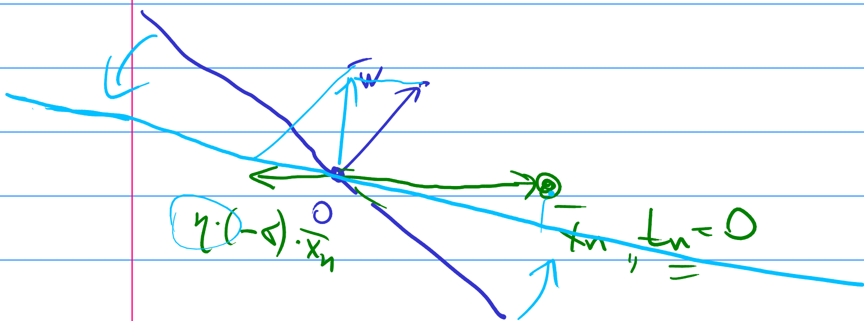
$$H_{\bar{w}}\left[\ln p(D|\bar{w})\right] = -\sum_n \sigma(\bar{w}^T\bar{x}_n)(1-\sigma(\bar{w}^T\bar{x}_n))\bar{x}_n\bar{x}_n^T$$

IRLS

$$= -\sum_n \boxed{\sigma_n(1-\sigma_n)}\,\bar{x}_n\bar{x}_n^T \qquad \sigma \approx 1$$

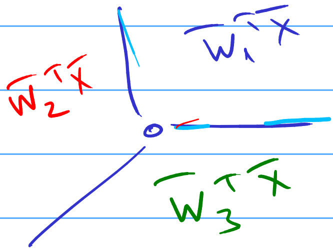






$$a_k = \ln p(C_k)p(\bar{x}|C_k) \approx \bar{w}_k^T\bar{x} \to \infty$$

$$p(C_1|\bar{x}) = \frac{p(\bar{x}|C_1)\,p(C_1)}{\sum_{k=1}^{K} p(\bar{x}|C_k)\,p(C_k)} = \frac{e^{a_1}}{\sum_{k=1}^{K} e^{a_k}} = \frac{e^{\bar{w}_1^T\bar{x}}}{\sum_{k=1}^{K} e^{\bar{w}_k^T\bar{x}}}$$



$$\mathrm{softmax}(a_1, \ldots, a_K) = \left(\frac{e^{a_1}}{\sum e^{a_k}}, \ \ldots, \ \frac{e^{a_K}}{\sum e^{a_\ell}}\right)$$

$$\bar{t}_n = (0 \ldots 1 \ldots 0) \qquad \text{one-hot repr.}$$

$$p(D|\bar{w}_1,\dots,\bar{w}_{K'}) = \prod_n p(\bar{t}_n|\bar{x}_n, W) =$$

$$= \prod_{n=1}^{N}\prod_{k=1}^{K'} p(t_{nk}|\bar{x}_n, W)^{t_{nk}}$$

$$\ln p(D|W) = \sum_n \sum_k t_{nk} \cdot \ln \overbrace{p(t_{nk}|\bar{x}_n, W)}^{= \frac{e^{\bar{w}_k^T \bar{x}_n}}{\sum e^{\bar{w}_\ell^T \bar{x}_n}}}$$