2022-09-20

Gaussian decision boundaries

$$p(x|C_1) = \frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma_1|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1)\right)$$

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

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

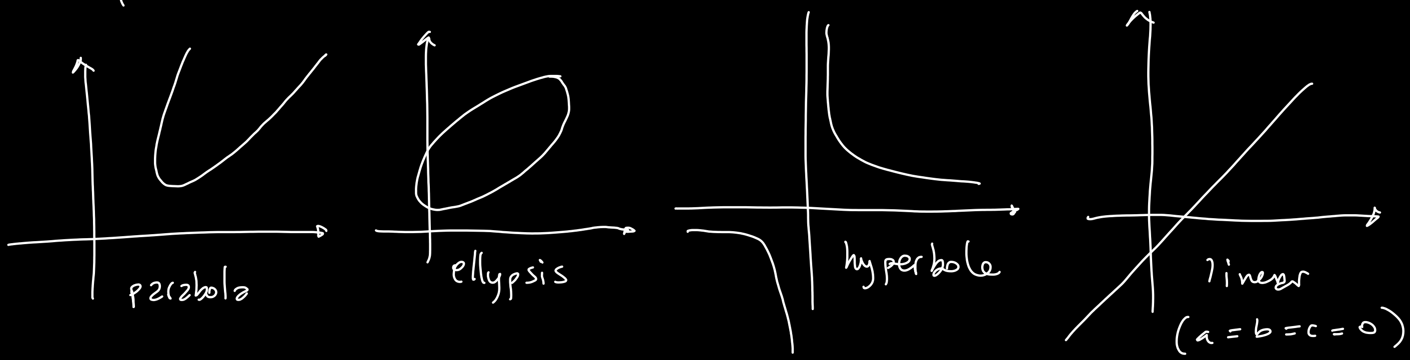
$$\ln p(C_1) \cancel{- \frac{D}{2}\ln 2\pi} - \frac{1}{2}\ln|\Sigma_1| - \frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) =$$

$$\ln p(C_2) \cancel{- \frac{D}{2}\ln 2\pi} - \frac{1}{2}\ln|\Sigma_2| - \frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)$$

$$\underbrace{\ln\frac{p(C_1)}{p(C_2)} - \frac{1}{2}\ln\frac{|\Sigma_1|}{|\Sigma_2|}}_{\text{const}} \underbrace{- \frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) + \frac{1}{2}(x-\mu_2)\Sigma_2^{-1}(x-\mu_2)}_{\{x_1^2,\ x_2^2,\ x_1x_2,\ x_1,\ x_2,\ \text{const}\}} = 0$$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$a x_1^2 + b x_2^2 + c\, x_1 x_2 + d x_1 + e x_2 + f = 0 \qquad \text{conic}$$



Special case $\Sigma_1 = \Sigma_2 = \Sigma$

log odds:

$$\ln\frac{p(C_1)}{p(C_2)} - \frac{1}{2}\left[(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) - (x-\mu_2)^T\Sigma^{-1}(x-\mu_2)\right]$$

$$\cancel{x^T\Sigma^{-1}x} - \underline{x^T\Sigma^{-1}\mu_1} - \underline{\mu_1^T\Sigma^{-1}x} + \mu_1^T\Sigma^{-1}\mu_1$$

$$\cancel{- x^T\Sigma^{-1}x} + \underline{x^T\Sigma^{-1}\mu_2} + \underline{\mu_2^T\Sigma^{-1}x} - \mu_2^T\Sigma^{-1}\mu_2$$

$$-2(\mu_1-\mu_2)^T\Sigma^{-1}x + \mu_1^T\Sigma^{-1}\mu_1 - \mu_2^T\Sigma^{-1}\mu_2$$

$$\underbrace{(\mu_1-\mu_2)^T\Sigma^{-1}}_{w^T} x \underbrace{- \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 + \ln\frac{P(C_1)}{P(C_2)}}_{w_0} = 0$$

$p(C_1|x) = \text{sig}(w^Tx)$ having defined $x = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_D \end{bmatrix}$ $w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_D \end{bmatrix}$

Discriminative training (Logistic Regression)

model: $t = \{0, 1\}$

posterior $t \sim \text{Bernoulli}(\lambda) = \lambda^t(1-\lambda)^{(1-t)}$

$$\lambda = \lambda(x) = \text{sig}(w^Tx) = \frac{1}{1+\exp(-w^Tx)}$$

Optimize posterior with ML

$$p(T|X,w) = \prod_{n=1}^{N} \lambda(x_n)^{t_n}(1-\lambda(x_n))^{(1-t_n)} \rightarrow \text{log domain}$$

$$\ln p(T|X,w) = \sum_{n=1}^{N} t_n \ln\lambda(x_n) + (1-t_n)\ln(1-\lambda(x_n))$$

$$= \sum_{n=1}^{N} t_n \ln \text{sig}(w^Tx_n) + (1-t_n)\ln(1-\text{sig}(w^Tx_n))$$

$$\text{sig}(w^Tx) = \frac{1}{1+e^{-w^Tx}}$$

$$\frac{d}{dw}\text{sig}(w^Tx) = -\frac{1}{(1+e^{-w^Tx})^2}\cdot\frac{d}{dw}(1+e^{-w^Tx}) = -\frac{1}{(1+e^{-w^Tx})^2}e^{-w^Tx}(-x) =$$

$$= \frac{1}{1+e^{-w^Tx}}\cdot\frac{e^{-w^Tx}}{1+e^{-w^Tx}}x = \text{sig}(w^Tx)\cdot\left[\frac{1+e^{-w^Tx}}{1+e^{-w^Tx}} - \frac{1}{1+e^{-w^Tx}}\right]x$$

$$= \text{sig}(w^Tx)(1-\text{sig}(w^Tx))x$$

Therefore

$$\frac{d}{dw}\ln\left(\mathrm{sig}(w^Tx)\right) = \frac{\mathrm{sig}(w^Tx)\left[1-\mathrm{sig}(w^Tx)\right]}{\mathrm{sig}\, w^Tx}\, x = \left(1-\mathrm{sig}(w^Tx)\right) x$$

$$\frac{d}{dw}\ln\left(1-\mathrm{sig}(w^Tx)\right) = \frac{-\mathrm{sig}(w^Tx)\left[1-\mathrm{sig}(w^Tx)\right]}{1-\mathrm{sig}(w^Tx)}\, x = -\mathrm{sig}(w^Tx)\, x$$

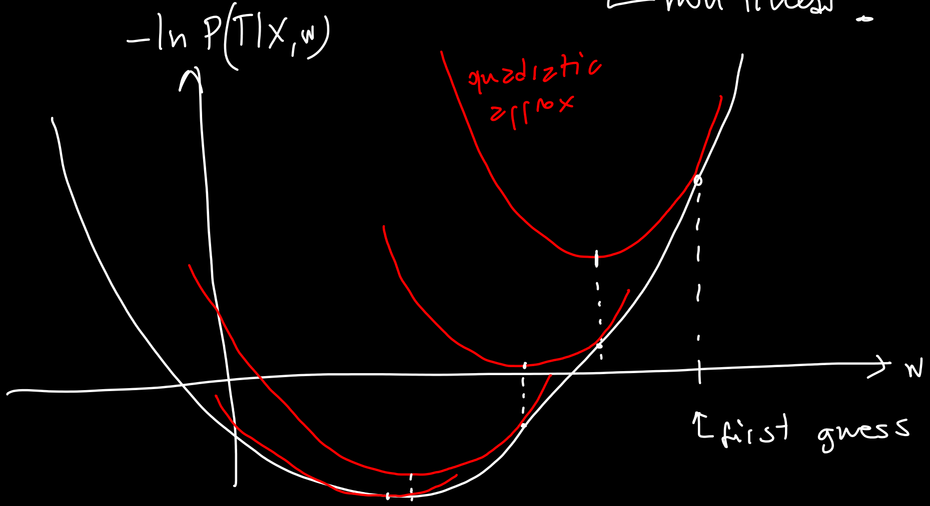
Maximum likelihood optimization:

$$\frac{d}{dw}\ln p(T|X,w) = \sum_{n=1}^{N} t_n\left(1-\mathrm{sig}(w^Tx)\right)x - (1-t_n)\,\mathrm{sig}(w^Tx)\,x =$$

$$= \sum_{n=1}^{N} t_n x - \cancel{t_n\,\mathrm{sig}(w^Tx)\,x} - \mathrm{sig}(w^Tx)\,x + \cancel{t_n\,\mathrm{sig}(w^Tx)\,x} =$$

$$= \sum_{n=1}^{N}\left[t_n - \mathrm{sig}(w^Tx)\right]x$$

↑ non linear! but convex!



gradient descend

or

Newton-Raphson