③ $\mathrm{Expo}(x|\lambda) = 1 - e^{-\lambda x} \quad x \geq 0$

$\hookrightarrow$ when $x \geq 0$

class prior is given

1) $p(y|x) \propto p(x|y=c)\, p(y=c)$

- $p(x|y=c) = \mathrm{Expo}(x|\lambda)$ with $\lambda_0 \neq \lambda_1$ known and fixed

- define $a = \log \dfrac{p(x|y=1)\,\cancel{p(y=1)}}{p(x|y=0)\,\cancel{p(y=0)}} = \log \dfrac{\lambda_1 e^{-\lambda_1 x}}{\lambda_0 e^{-\lambda_0 x}} = \log\left(\dfrac{\lambda_1}{\lambda_0}\right) + \log\left(e^{-\lambda_1 x + \lambda_0 x}\right)$

$$= \underbrace{\log\left(\frac{\lambda_1}{\lambda_0}\right) + x(\lambda_0 - \lambda_1)}_{\text{linear function}}$$

from class

- for two classes, $p(y=1|x) = 1/(1+\exp(-a)) = 1/\left(1+\exp\left(-\left(\log\left(\frac{\lambda_1}{\lambda_0}\right) + x(\lambda_0 - \lambda_1)\right)\right)\right)$

$= 1/\left(1+\exp(-(w_0 + wx))\right)$, with $w_0 = \log\left(\frac{\lambda_1}{\lambda_0}\right)$

$:= \sigma(wx + w_0)$ $\quad w = (\lambda_0 - \lambda_1)$

$\Rightarrow$ thus, since we only observe x once, this is a bernoulli experiment $\sim \mathrm{Bernoulli}(x, \sigma(wx + w_0))$

2) $\hat{y} = \arg\max_k p(y=k|x)$

$\Rightarrow p(y=1|x) \Leftrightarrow \sigma(wx+w_0) > 1/2$

$\hookrightarrow p(y=1|x) = 1 - p(y=0|x)$

$\Rightarrow$ if $p(y=1|x) > p(y=0|x) = 1 - p(y=1|x)$

$\Leftrightarrow p(y=1|x) > 1/2$

$\Leftrightarrow 1/\left(1+\exp\left(-\log\left(\frac{\lambda_1}{\lambda_0}\right) - x(\lambda_0 - \lambda_1)\right)\right) > 1/2 \Leftrightarrow$

$\Leftrightarrow 2 > 1 + \exp\left(-\log\left(\frac{\lambda_1}{\lambda_0}\right) - x(\lambda_0 - \lambda_1)\right) \Leftrightarrow$

$\Leftrightarrow 1 > \exp\left(-\log\left(\frac{\lambda_1}{\lambda_0}\right) - x(\lambda_0 - \lambda_1)\right) \Leftrightarrow$

$\Leftrightarrow -\log\left(\frac{\lambda_1}{\lambda_0}\right) - x(\lambda_0 - \lambda_1) > 0 \Leftrightarrow$

$\Leftrightarrow x < \dfrac{-\log(\lambda_1/\lambda_0)}{\lambda_0 - \lambda_1}$ //

④ $p(y \mid w, X) = \prod_{i=1}^{N} p(y_i \mid x_i, w)$

$= \prod_{i=1}^{N} \sigma(w^T x_i)^{y_i} (1 - \sigma(w^T x_i))^{1-y_i}$

$\Rightarrow$ mle is identical to solving $\arg\min_w -\log\left(\prod_{i=1}^{N} \sigma(w^T x_i)^{y_i} (1-\sigma(w^T x_i))^{1-y_i}\right)$

$\cdot \ -\log\left(\prod_{i=1}^{N} \sigma(w^T x_i)^{y_i}(1-\sigma(w^T x_i))^{1-y_i}\right) = -\left(\sum_{i=1}^{N} \log\left(\sigma(w^T x_i)^{y_i} (1-\sigma(w^T x_i))^{1-y_i}\right)\right) = f(w, x)$

$\Rightarrow \frac{\partial f}{\partial w} = -\left(\sum_{i=1}^{N} y_i \frac{\sigma'(w^T x_i)}{\sigma(w^T x_i)} + (1-y_i)\frac{(-1)\sigma'(w^T x_i)}{1-\sigma(w^T x_i)}\right)$

$= -\left(\sum_{i=1}^{N} y_i \left(\cancel{x e^{-w^T x}} \cdot \sigma\right) - (1-y_i)\left(\frac{\cancel{x e^{-w^T x}} \cdot \sigma^2}{1-\sigma}\right)\right)$

less clutter

$\frac{\partial \sigma}{\partial w} = \frac{\partial\left(1/(1+\exp(-wx))\right)}{\partial w} =$ $\sim (1+\exp(-wx))^{-1}$

$= -1(-x e^{-wx})(1+\exp(-wx))^{-2}$

$= \frac{x e^{-wx}}{(1+\exp(-wx))} = x e^{-wx} \cdot \sigma^2$

$\Rightarrow \frac{\partial f}{\partial w} = 0 \Leftrightarrow \sum_{i=1}^{N} y_i = \sum_{i=1}^{N} (1-y_i) \frac{\sigma}{1-\sigma}$ $\Leftrightarrow$

$\Leftrightarrow \sum_{i=1}^{N} y_i = N - \frac{\sigma}{1-\sigma} \sum_{i=1}^{N} y_i$ $\Leftrightarrow$

$\Leftrightarrow m_1 = N - \frac{\sigma}{1-\sigma} m_1$ $\Leftrightarrow$

$\Leftrightarrow \left(1 + \frac{\sigma}{1-\sigma}\right) m_1 = N$ $\Leftrightarrow$

≤ 1

$\Leftrightarrow \frac{1}{N} \sum_{i=1}^{N} y_i = 1/(1-\sigma)$

$\Rightarrow 1/(1-\sigma) \leq 1$

$1/(1-\sigma) \leq 1 \Leftrightarrow$

$\Leftrightarrow 1 \leq 1 - \frac{1}{1+e^{-w^T x}}$

$\Leftrightarrow 0 \leq (-1)\frac{1}{1+e^{-w^T x}}$

$\Rightarrow -1/(1+e^{-w^T x}) \to 0$

$\Leftrightarrow (1+e^{-w^T x}) \to \infty$

$\Leftrightarrow e^{w^T x} \to 0$

$\Leftrightarrow w^T x \to \infty$

$\Leftrightarrow \|w\| \to \infty$ //

instead of using the normal negative log likelihood, we can add a regularization factor $\lambda \|w\|$ so that an optimum solution won't have $\|w\| \to \infty$

$$E(\boldsymbol{w}) = -\log p(\boldsymbol{y} \mid \boldsymbol{w}, \boldsymbol{X}) + \lambda \|\boldsymbol{w}\|_q^q \quad (35)$$

(5) $p(y=c_1|x) = \sigma(w_1^T x + w_{0_1})$

$$= \frac{\exp(w_1^T x + w_{0_1})}{\exp(w_1^T x + w_{0_1}) + \exp(w_2^T x + w_{0_2})} = \frac{\exp(w_1^T x + w_{0_1})}{\exp(w_1^T x + w_{0_1})} \times \frac{1}{1 + \exp(w_2^T x + w_{0_2}) \cdot \exp(-w_1^T x + w_{0_1})}$$

$$= \frac{1}{1 + \exp\left((w_2^T - w_1^T)x + (w_{0_2} - w_{0_1})\right)}$$

$$= \frac{1}{1 + \exp(-a)} = \sigma(w^T x - w_0) \quad //$$

in the softmax setting a = decision boundary

$\Rightarrow w_1^T x - w_{0_1} = w_2^T x - w_{0_2} = a$

$\rightarrow \underbrace{(w_1^T - w_2^T)}_{w^T} x - \underbrace{(w_{0_1} - w_{0_2})}_{w_0} = a$

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

$$= (1+e^{-a})^{-1}\left[e^{-a} \cdot (1+e^{-a})^{-1}\right]$$

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

$$= \sigma(a)\left(\sigma(a) \cdot e^{-a}\right) =$$

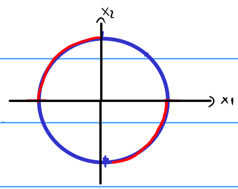
$e^{-a} = 1 + e^{-a} - 1 = \sigma(a)^{-1} - 1$

$$= \sigma(a)\left(\sigma(a) \cdot (\sigma(a)^{-1} - 1)\right)$$

$$= \sigma(a)(1 - \sigma(a)) \quad //$$

(7) $\phi(x_1, x_2) = \text{sign}(tg\,\theta) = \text{sign}(y/x)$

$\rightarrow$ assuming $x \neq 0$



where θ = angle(x_1, x_2)

(8) • $p(y=1|x) = p(y=0|x)$

if we compute a and equate it to 0 we solve the problem, $a = \log \frac{p(x|y=1)\,p(y=1)}{p(x|y=0)\,p(y=0)} = 0 \Leftrightarrow$

$\Leftrightarrow \frac{p(x|y=1)\,p(y=1)}{p(x|y=0)\,p(y=0)} = 1 \Leftrightarrow p(x|y=1)\,p(y=1) = p(x|y=0)\,p(y=0) \Leftrightarrow p(y=1|x) = p(y=0|x)$ //

$$\rightarrow a = \log \frac{p(x|y=1)\,p(y=1)}{p(x|y=0)\,p(y=0)} = \log \frac{N(x|\mu_1, \Sigma_1)\,\pi_1}{N(x|\mu_0, \Sigma_0)\,\pi_0} = \log \left[\frac{\exp\left(-1/2\,(x-\mu_1)^T \Sigma_1^{-1} (x-\mu_1)\right)}{\exp\left(-1/2\,(x-\mu_0)^T \Sigma_0^{-1} (x-\mu_0)\right)} \sqrt{\frac{|\Sigma_0|}{|\Sigma_1|}} \frac{\pi_1}{\pi_0} \right]$$

$$= -1/2\,(x-\mu_1)^T \Sigma_1^{-1} (x-\mu_1) - 1/2\,(x-\mu_0) \Sigma_0^{-1} (x-\mu_0) + 1/2 \log\left(\frac{|\Sigma_0|}{|\Sigma_1|}\right) + \log(\pi_1/\pi_0) =$$

$$= \underbrace{1/2\, x^T (\Sigma_0^{-1} - \Sigma_1^{-1})\, x}_{x^T A x} + \underbrace{x^T [\Sigma_1^{-1}\mu_1 - \Sigma_0^{-1}\mu_0]}_{b^T x} \underbrace{- 1/2 \cdot \mu_1^T \Sigma_1^{-1} \mu_1 - 1/2 \cdot \mu_0^T \Sigma_0^{-1} \mu_0 + 1/2 \cdot \log\left(\frac{|\Sigma_0|}{|\Sigma_1|}\right) + \log(\pi_1/\pi_0)}_{c}$$

$\Rightarrow A = 1/2\,(\Sigma_0^{-1} - \Sigma_1^{-1})$

$\Rightarrow b = \Sigma_1^{-1}\mu_1 - \Sigma_0^{-1}\mu_0$

$\Rightarrow c = -1/2 \cdot \mu_1^T \Sigma_1^{-1} \mu_1 - 1/2 \cdot \mu_0^T \Sigma_0^{-1} \mu_0 + 1/2 \cdot \log\left(\frac{|\Sigma_0|}{|\Sigma_1|}\right) + \log(\pi_1/\pi_0)$