P1

Wednesday, February 2, 2022 4:32 PM

I discussed this hw with William Plucknett (wbpluck) and David Abraham (dpabra)

a) $$\ell(\vec{w}) = \sum_{i=1}^{N} y^{(i)} \log h(\vec{x}^{(i)}) + (1-y^{(i)}) \log(1 - h(\vec{x}^{(i)}))$$

$$h(\vec{x}) = \sigma(\vec{w}^T \vec{x}^{(i)}) = \frac{1}{1+\exp(-w^T x)} = \sigma^{(i)}$$

$$\nabla_{\vec{w}} \ell(\vec{w}) = \sum_{i=1}^{N} \nabla_{\vec{w}} \; y^{(i)} \log h(\vec{x}^{(i)}) + (1-y^{(i)}) \log(1 - h(\vec{x}^{(i)}))$$

$$= \sum_{i=1}^{N} \left(y^{(i)} (1-\sigma^{(i)}) - (1-y^{(i)}) \sigma^{(i)} \right) \nabla_{\vec{w}} (\vec{w}^T \vec{x}^{(i)})$$

$$= \sum_{i=1}^{N} (y^{(i)} - \sigma^{(i)}) \vec{x}^{(i)}$$

$$H = \frac{\partial^2 \ell(\vec{w})}{\partial \vec{w}_i \, \partial \vec{w}_j} \Rightarrow \frac{\partial}{\partial \vec{w}} \nabla_{\vec{w}} \ell(\vec{w}) = \frac{\partial}{\partial \vec{w}} \sum_{i=1}^{N} (y^{(i)} - \sigma^{(i)}) \vec{x}^{(i)}$$

$$= -\sum_{i=1}^{N} \frac{\partial}{\partial \vec{w}} \sigma^{(i)} \vec{x}^{(i)}$$

$$= -\sum_{i=1}^{N} \frac{\partial}{\partial \vec{w}} \frac{1}{1+\exp(\vec{w}^T \vec{x}^{(i)})} \vec{x}^{(i)}$$

$$= \sum_{i=1}^{N} \underbrace{-\sigma(\vec{w}^T \vec{x}^{(i)})(1 - \sigma(\vec{w}^T \vec{x}^{(i)}))}_{d} \vec{x}^{(i)} \vec{x}^{(i)}$$

$$\boxed{H = X^T D X}$$ where $D_{ii} = d = -\sigma^{(i)}(1-\sigma^{(i)})$

By definition of Gram matrix $X^T X \succeq 0$

& $\because$ $0 < \sigma^{(i)} \leq 1$ $\quad -\sigma^{(i)} < 0 \Rightarrow d < 0$

$\Rightarrow X^T D X$ is negative semidefinite $\Rightarrow H$ is concave

For any $\vec{z}$

$$\vec{z}^T H \vec{z} = \vec{z}^T X^T D X \vec{z}$$

$$= \Sigma_i \Sigma_j z_i x_{ii} \, d \, x_j z_j$$

$$= \Sigma_i \Sigma_j \, d \, z_i x_i x_j z_j$$

$$= \Sigma_i \, d \, (\vec{x}^{(i)T} \vec{z})^2$$

$d < 0$ as shown above & $(\vec{x}^{(i)T} \vec{z})^2 \geq 0$

$\Rightarrow \Sigma_i \, d \, (\vec{x}^{(i)T} \vec{z})^2 \leq 0$

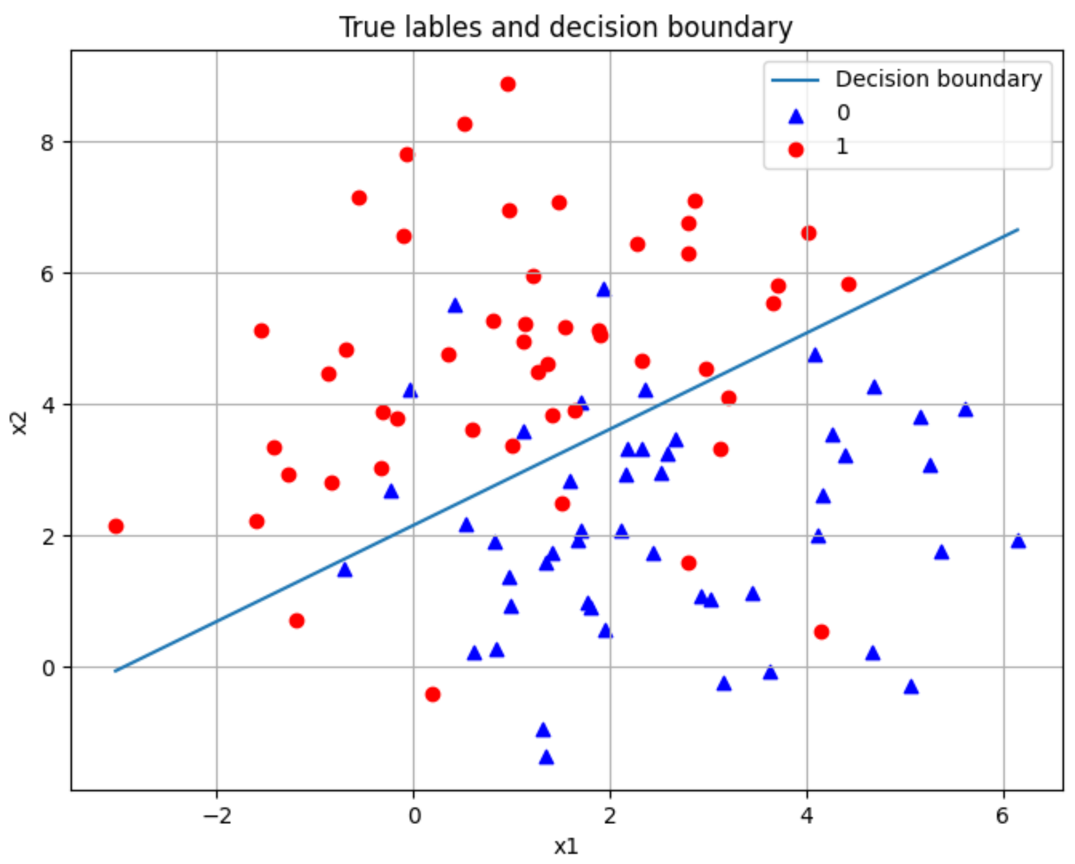
$$\boxed{\Rightarrow \vec{z}^T H \vec{z} \leq 0}$$ <u>Hence proven!</u>

(b) The $\vec{w}$ update under Newton-Raphson is given by $\vec{w} := \vec{w} - H^{-1} \nabla_w E$

where $H_{ij}(\vec{w}) = \dfrac{\partial^2 E(\vec{w})}{\partial w_i w_j}$

The trained weights from the newton method $\mathbf{w}$ = [[-1.84922892]
[-0.62814188]
[0.85846843]]

(c)



P2

Wednesday, February 2, 2022 4:32 PM

I discussed this hw with William Plucknett (wbpluck) and David Abraham (dpabra)

(a)

Top: $\nabla_{\vec{w}_m} \ell(\vec{w}) = \sum_{i=1}^{N} \phi(\vec{x}^{(i)}) \left[\mathbb{I}(y^{(i)}=m) - \frac{\exp(\vec{w}_m^T \phi(\vec{x}^{(i)}))}{1+\sum_{j=1}^{K-1} \exp(\vec{w}_j^T \phi(\vec{x}^{(i)}))} \right]$

$$\ell(\vec{w}) = \sum_{i=1}^{N} \sum_{k=1}^{K} \log\left(\left[p(y^{(i)}=k \mid \vec{x}^{(i)}, \vec{w}) \right]^{\mathbb{I}(y^{(i)}=k)} \right)$$

$$\ell(\vec{w}) = \sum_{i=1}^{N} \sum_{k=1}^{K} \mathbb{I}(y^{(i)}=k) \log\left(p(y^{(i)}=k \mid \vec{x}^{(i)}, \vec{w}) \right)$$

$$\ell(\vec{w}) = \sum_{i=1}^{N} \left(\sum_{k=1}^{K-1} \left(\mathbb{I}(y^{(i)}=k) \log\left[p(y^{(i)}=k \mid \vec{x}^{(i)}, \vec{w}) \right] \right) + \mathbb{I}(y^{(i)}=K) \log\left[p(y^{(i)}=K \mid \vec{x}^{(i)}, \vec{w}) \right] \right)$$

$$\ell(\vec{w}) = \sum_{i=1}^{N} \sum_{k=1}^{K-1} \mathbb{I}(y^{(i)}=k) \left[\log\left[\exp(\vec{w}_k^T \phi(\vec{x}^{(i)})) \right] - \log\left(1+\sum_{j=1}^{K-1} \exp(\vec{w}_j^T \phi(\vec{x}^{(i)}))\right) \right] + \mathbb{I}(y^{(i)}=K) \left[\log(1) - \log\left(1+\sum_{j=1}^{K-1} \exp(\vec{w}_j^T \phi(\vec{x}^{(i)}))\right) \right]$$

$$\begin{aligned}
\ell(\vec{w}) = \sum_{i=1}^{N} \Big\{ & \mathbb{I}(y^{(i)}=1) \left[\vec{w}_1^T \phi(\vec{x}^{(i)}) - \log\left(1+\sum_{j=1}^{K-1} \exp(\vec{w}_j^T \phi(\vec{x}^{(i)}))\right) \right] \\
& \vdots \\
& + \mathbb{I}(y^{(i)}=m) \left[\vec{w}_m^T \phi(\vec{x}^{(i)}) - \log\left(1+\sum_{j=1}^{K-1} \exp(\vec{w}_j^T \phi(\vec{x}^{(i)}))\right) \right] \\
& \vdots \\
& + \mathbb{I}(y^{(i)}=K-1) \left[\vec{w}_{K-1}^T \phi(\vec{x}^{(i)}) - \log\left(1+\sum_{j=1}^{K-1} \exp(\vec{w}_j^T \phi(\vec{x}^{(i)}))\right) \right] \\
& + \mathbb{I}(y^{(i)}=K) \left[-\log\left(1+\sum_{j=1}^{K-1} \exp(\vec{w}_j^T \phi(\vec{x}^{(i)}))\right) \right] \Big\}
\end{aligned}$$

using the given hint and splitting into 2 cases

Case I: $y^{(i)} = k = m$

$$\nabla_{\vec{w}_m} \ell(w) = \sum_{i=1}^{N} \left(\mathbb{I}\{y^{(i)}=k=m\} \phi(\vec{x}^{(i)}) - \mathbb{I}\{y^{(i)}=k=m\} \frac{\boxed{\sum_{j=1}^{K-1} \exp(w_j^T \phi(\vec{x}^{(i)})) \nabla_{\vec{w}_m} (\vec{w}_j^T \phi(\vec{x}^{(i)}))}}{1+\sum_{j=1}^{K-1} \exp(\vec{w}_j^T \phi(\vec{x}^{(i)}))} \right)$$

terms become 0 $\because \nabla_{\vec{w}_m} (\vec{w}_j^T \phi(\vec{x}^{(i)})) = 0 \quad \forall j \neq m$

$$= \sum_{i=1}^{N} \left(\mathbb{I}\{y^{(i)}=m\} \phi(\vec{x}^{(i)}) \right) - \mathbb{I}\{y^{(i)}=m\} \frac{\exp(\vec{w}_m^T \phi(\vec{x}^{(i)})) \phi(\vec{x}^{(i)})}{1+\sum_{j=1}^{K-1} \exp(\vec{w}_j^T \phi(\vec{x}^{(i)}))}$$

Case II: $y^{(i)} = k \neq m$

$$\ell(\vec{w}) = \sum_{i=1}^{N} \sum_{k: k\neq m} \mathbb{I}\{y^{(i)}=k\} \left[\vec{w}_k^T \phi(\vec{x}^{(i)}) - \log\left(1+\sum_{j=1}^{K-1} w_j^T \phi(\vec{x}^{(i)})\right) \right]$$

$y^{(i)}$ only has a single value, thus can only equal one k in the inner summation. For this value of k $\mathbb{I}\{y^{(i)}=k\}=1$

$$\Rightarrow \nabla_{\vec{w}_m} \ell(\vec{w}) = \sum_{i=1}^{N} \frac{-\exp(\vec{w}_m^T \phi(\vec{x}^{(i)})) \phi(\vec{x}^{(i)})}{1+\sum_{j=1}^{K-1} \vec{w}_j^T \phi(\vec{x}^{(i)})}$$

This shows $\nabla_{\vec{w}_m} \ell(\vec{w}) = \begin{cases} \sum_{i=1}^{N} \left(\mathbb{I}\{y^{(i)}=m\} \phi(\vec{x}^{(i)}) \right) - \mathbb{I}\{y^{(i)}=m\} p(y^{(i)}=m \mid \vec{x}^{(i)}, \vec{w}) \phi(\vec{x}^{(i)}) \\ \sum_{i=1}^{N} -p(y^{(i)}=m \mid \vec{x}^{(i)}, \vec{w}) \phi(\vec{x}^{(i)}) \end{cases}$

$$\boxed{\Rightarrow \nabla_{\vec{w}_m} \ell(\vec{w}) = \sum_{i=1}^{N} \phi(\vec{x}^{(i)}) \left[\mathbb{I}\{y^{(i)}=m\} - p(y^{(i)}=m \mid \vec{x}^{(i)}, \vec{w}) \right]}$$

(b) Accuracy: 94%

P3a

Sunday, February 6, 2022 2:48 PM

I discussed this hw with William Plucknett (wbpluck) and David Abraham (dpabra)

a) $$p(y=1 \mid \vec{x}) = \frac{p(\vec{x} \mid y=1)\, p(y=1)}{p(\vec{x} \mid y=1)\, p(y=1) + p(\vec{x} \mid y=0)\, p(y=0)}$$

$$= \frac{1}{1 + \frac{p(\vec{x} \mid y=0)\, p(y=0)}{p(\vec{x} \mid y=1)\, p(y=1)}} \quad - ①$$

$$\frac{p(\vec{x} \mid y=0)\, p(y=0)}{p(\vec{x} \mid y=1)\, p(y=1)} = \frac{e^{-\frac{1}{2}(\vec{x}-\mu_0)^T \Sigma^{-1} (\vec{x}-\mu_0)}}{e^{-\frac{1}{2}(\vec{x}-\mu_1)^T \Sigma^{-1} (\vec{x}-\mu_1)}} \cdot \frac{1-\phi}{\phi}$$

$$= \left(\frac{1-\phi}{\phi}\right) \frac{\exp\left(-\frac{1}{2}\vec{x}^T\Sigma^{-1}\vec{x} + \frac{1}{2}\mu_0^T\Sigma^{-1}\vec{x} + \frac{1}{2}\vec{x}^T\Sigma^{-1}\mu_0 - \frac{1}{2}\mu_0^T\Sigma^{-1}\mu_0\right)}{\exp\left(-\frac{1}{2}\vec{x}^T\Sigma^{-1}\vec{x} + \frac{1}{2}\mu_1^T\Sigma^{-1}\vec{x} + \frac{1}{2}\vec{x}^T\Sigma^{-1}\mu_1 - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1\right)}$$

$$= \frac{(1-\phi)}{\phi} \exp\left(\frac{1}{2}\mu_0^T\Sigma^{-1}\vec{x} + \frac{1}{2}x^T\Sigma^{-1}\mu_0 - \frac{1}{2}\mu_0^T\Sigma^{-1}\mu_0 - \frac{1}{2}\mu_1^T\Sigma^{-1}\vec{x} - \frac{1}{2}\vec{x}^T\Sigma^{-1}\mu_1 + \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1\right)$$

$$= \frac{(1-\phi)}{\phi} \exp\left(\frac{1}{2}\mu_0^T\Sigma^{-1}(\vec{x}-\mu_0) - \frac{1}{2}\mu_1^T\Sigma^{-1}(\vec{x}-\mu_1) + \frac{1}{2}\vec{x}^T\Sigma^{-1}\mu_0 - \frac{1}{2}\vec{x}^T\Sigma^{-1}\mu_1\right)$$

$$= \frac{(1-\phi)}{\phi} \exp\left(\frac{1}{2}\mu_0^T\Sigma^{-1}(\vec{x}-\mu_0) - \frac{1}{2}\mu_1^T\Sigma^{-1}(\vec{x}-\mu_1) + \frac{1}{2}\vec{x}^T\Sigma^{-1}(\mu_0-\mu_1)\right)$$

plugging back in ①

$$p(y=1 \mid \vec{x}) = \frac{1}{1 + \frac{(1-\phi)}{\phi} \exp\left(\frac{1}{2}\mu_0^T\Sigma^{-1}(\vec{x}-\mu_0) - \frac{1}{2}\mu_1^T\Sigma^{-1}(\vec{x}-\mu_1) + \frac{1}{2}x^T\Sigma^{-1}(\mu_0-\mu_1)\right)}$$

$$= \frac{1}{1 + \exp\left(\log\left(\frac{1-\phi}{\phi}\right) + \frac{1}{2}\mu_0^T\Sigma^{-1}(\vec{x}-\mu_0) - \frac{1}{2}\mu_1^T\Sigma^{-1}(\vec{x}-\mu_1) - \frac{1}{2}x^T\underbrace{\Sigma^{-1}(\mu_1-\mu_0)}_{\vec{w}_\alpha}\right)}$$

$$= \frac{1}{1 + \exp\left(\log\left(\frac{1-\phi}{\phi}\right) - \frac{1}{2}\mu_0^T\Sigma^{-1}\mu_0 + \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \frac{1}{2}\mu_0^T\Sigma^{-1}\vec{x} - \frac{1}{2}\mu_1^T\Sigma^{-1}\vec{x} - \frac{1}{2}\vec{x}^T\vec{w}_\alpha\right)}$$

$$= \frac{1}{1 + \exp\left(\log\left(\frac{1-\phi}{\phi}\right) - \frac{1}{2}\mu_0^T\Sigma^{-1}\mu_0 + \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 - \frac{1}{2}(\mu_1^T - \mu_0^T)\Sigma^{-1}\vec{x} - \frac{1}{2}\vec{w}_\alpha^T\vec{x}\right)}$$

$\because \vec{x}^T\vec{w}_\alpha$ is a scalar $\vec{x}^T\vec{w}_\alpha = \vec{w}_\alpha^T\vec{x}$

$\because \Sigma$ is a covariance matrix $\Sigma^{-1} = (\Sigma^{-1})^T$
$\Rightarrow (\mu_1^T - \mu_0^T)\Sigma^{-1} = (\Sigma^{-1}(\mu_1-\mu_0))^T = \vec{w}_\alpha^T$

$$= \frac{1}{1 + \exp\left(\underbrace{-\left(\log\left(\frac{\phi}{1-\phi}\right) + \frac{1}{2}\mu_0^T\Sigma^{-1}\mu_0 - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1\right)}_{w_0} - \frac{1}{2}\underbrace{(\mu_1^T - \mu_0^T)\Sigma^{-1}}_{\vec{w}_\alpha^T}\vec{x} - \frac{1}{2}\vec{w}_\alpha^T\vec{x}\right)}$$

$$= \frac{1}{1 + \exp(-w_0 - \vec{w}_\alpha^T\vec{x})}$$

$$p(y=1 \mid \vec{x}) = \frac{1}{1 + \exp(-\vec{w}^T\vec{x})}$$

where $\vec{w} = \begin{bmatrix} w_0 \\ \vec{w}_\alpha \end{bmatrix}$ $\quad w_0 = \log\left(\frac{\phi}{1-\phi}\right) + \frac{1}{2}\mu_0^T\Sigma^{-1}\mu_0 - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1$

$\vec{w}_\alpha = \Sigma^{-1}(\mu_1 - \mu_0)$

& $\vec{x}$ is redefined to be M+1 dimensional vectors by adding an extra co-ordinate $x_0 = 1$

P3b & P3c

Sunday, February 6, 2022 6:31 PM

I discussed this hw with William Plucknett (wbpluck) and David Abraham (dpabra)

3(c) let n_1 be the points where $y^{(i)}=1$

then $l(\phi, \mu_0, \mu_1, \Sigma) = \log\left(\prod_{i=1}^{n_1} p(x^{(i)} | y^{(i)}=1) p(y^{(i)}=1) \mathbb{I}\{y^{(i)}=1\}\right)$

$+ \log\left(\prod_{i=n_1+1}^{N} p(x^{(i)} | y^{(i)}=0) p(y^{(i)}=0) \mathbb{I}\{y^{(i)}=0\}\right)$

$= \log\left(\prod_{i=1}^{n_1} p(x^{(i)} | y^{(i)}=1)\, \phi\, \mathbb{I}\{y^{(i)}=1\}\right) + \log\left(\prod_{i=n_1+1}^{N} p(x^{(i)} | y^{(i)}=0)(1-\phi)\, \mathbb{I}\{y^{(i)}=0\}\right)$

$= \log\left(\prod_{i=1}^{n_1} p(x^{(i)} | y^{(i)}=1)\right) + \log\left(\prod_{i=1}^{n_1} \phi\, \mathbb{I}\{y^{(i)}=1\}\right) + \log\left(\prod_{i=n_1+1}^{N} p(x^{(i)} | y^{(i)}=0)\right) + \log\left(\prod_{i=n_1+1}^{N} (1-\phi)\mathbb{I}\{y^{(i)}=0\}\right)$

$= \sum_{i=1}^{n_1} \log\left(\frac{1}{(2\pi)^{M/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x^{(i)}-\mu_1)^T \Sigma^{-1} (x^{(i)}-\mu_1)\right)\right) + \sum_{i=1}^{n_1} \log(\phi\, \mathbb{I}\{y^{(i)}=1\})$

$+ \sum_{i=n_1+1}^{N} \log\left(\frac{1}{(2\pi)^{M/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x^{(i)}-\mu_0)^T \Sigma^{-1} (x^{(i)}-\mu_0)\right)\right) + \sum_{i=n_1+1}^{N} \log((1-\phi)\mathbb{I}\{y^{(i)}=0\})$

$l = \sum_{i=1}^{n_1}\left(-\frac{1}{2}(x^{(i)}-\mu_1)^T \Sigma^{-1} (x^{(i)}-\mu_1) - \frac{M}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma|\right) + \sum_{i=1}^{n_1} \log(\phi\, \mathbb{I}\{y^{(i)}=1\})$

$+ \sum_{i=n_1+1}^{N}\left(-\frac{1}{2}(x^{(i)}-\mu_0)^T \Sigma^{-1} (x^{(i)}-\mu_0) - \frac{M}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma|\right) + \sum_{i=n_1+1}^{N} \log((1-\phi)\mathbb{I}\{y^{(i)}=0\})$

ML estimate ϕ:

$$\frac{\partial l}{\partial \phi} = \sum_{i=1}^{n_1} \frac{\mathbb{I}\{y^{(i)}=1\}}{\phi} - \sum_{i=n_1+1}^{N} \frac{\mathbb{I}\{y^{(i)}=0\}}{1-\phi}$$

set $\frac{\partial l}{\partial \phi}=0 \Rightarrow$ $0 = \sum_{i=1}^{n_1} \frac{(1-\phi)\mathbb{I}\{y^{(i)}=1\}}{\phi(1-\phi)} - \sum_{i=n_1+1}^{N} \frac{\phi\, \mathbb{I}\{y^{(i)}=0\}}{\phi(1-\phi)}$

$$0 = \sum_{i=1}^{n_1} (1-\phi)\mathbb{I}\{y^{(i)}=1\} - \sum_{i=n_1+1}^{N} \phi\, \mathbb{I}\{y^{(i)}=0\}$$

$$0 = \sum_{i=1}^{n_1} \mathbb{I}\{y^{(i)}=1\} - \sum_{i=1}^{n_1} \phi\, \mathbb{I}\{y^{(i)}=1\} - \sum_{i=n_1+1}^{N} \phi\, \mathbb{I}\{y^{(i)}=0\}$$

$$\Rightarrow N\phi = \sum_{i=1}^{n_1} \mathbb{I}\{y^{(i)}=1\}$$

$$\rightarrow \phi = \frac{1}{N}\sum_{i=1}^{n_1} \mathbb{I}\{y^{(i)}=1\}$$

$$= \frac{1}{N}\left(\sum_{i=1}^{n_1} \mathbb{I}\{y^{(i)}=1\} + \underbrace{\sum_{i=n_1+1}^{N} \mathbb{I}\{y^{(i)}=1\}}\right)$$

we know this term is always 0 $\because$ $y^{(i)}=0$ in n_1+1 to N

$$\Rightarrow \phi = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\{y^{(i)}=1\}$$

ML estimate μ_1

$\frac{\partial l}{\partial \mu_1} = +\frac{1}{2}\sum_{i=1}^{n_1} 2(x^{(i)}-\mu_1)\Sigma^{-1}$, set $\frac{\partial l}{\partial \mu_1}=0$

$$\Rightarrow 0 = \sum_{i=1}^{n_1} (x^{(i)}-\mu_1)\Sigma^{-1}$$

$$\Rightarrow 0 = \sum_{i=1}^{N} \mathbb{I}\{y^{(i)}=1\} x^{(i)}\Sigma^{-1} - \sum_{i=1}^{N} \mathbb{I}\{y^{(i)}=1\}\mu_1\Sigma^{-1}$$

$$\sum_{i=1}^{N} \mathbb{I}\{y^{(i)}=1\}\mu_1\Sigma^{-1} = \sum_{i=1}^{N} \mathbb{I}\{y^{(i)}=1\} x^{(i)}\Sigma^{-1}$$

$$\Rightarrow \mu_1 = \frac{\sum_{i=1}^{N} \mathbb{I}\{y^{(i)}=1\} x^{(i)}}{\sum_{i=1}^{N} \mathbb{I}\{y^{(i)}=1\}}$$

ML estimate for μ_0: Similar to μ_1 but for $y^{(i)}=0$

$$\Rightarrow \mu_0 = \frac{\sum_{i=1}^{N} \mathbb{I}\{y^{(i)}=0\} x^{(i)}}{\sum_{i=1}^{N} \mathbb{I}\{y^{(i)}=0\}}$$

ML estimate for Σ:

$$\frac{\partial l}{\partial \Sigma^{-1}} = 0 = -\frac{1}{2}\sum_{i=1}^{N}\left((x^{(i)}-\mu_{y^{(i)}})(x^{(i)}-\mu_{y^{(i)}})^T - \nabla_{\Sigma^{-1}}\log|\Sigma^{-1}|\right)$$

$$\sum_{i=1}^{N} \nabla_{\Sigma^{-1}}\log|\Sigma^{-1}| = \sum_{i=1}^{N}(x^{(i)}-\mu_{y^{(i)}})(x^{(i)}-\mu_{y^{(i)}})^T$$

$$N\Sigma = \sum_{i=1}^{N}(x^{(i)}-\mu_{y^{(i)}})(x^{(i)}-\mu_{y^{(i)}})^T$$

$$\Sigma = \frac{1}{N}\sum_{i=1}^{N}(x^{(i)}-\mu_{y^{(i)}})(x^{(i)}-\mu_{y^{(i)}})^T$$

3(b) As shown in 3(c) the ML estimates work for all values of M including M = 1.
3(b) is just a special case of 3(c)

P4

Tuesday, February 8, 2022 4:55 PM

I discussed this hw with William Plucknett (wbpluck) and David Abraham (dpabra)

A) Error: 1.6250%

B) Top 5 spam words : ['httpaddr' 'spam' 'unsubscrib' 'ebai' 'valet']

C) Training size: 50
Error: 3.8750%

Training size: 100
Error: 2.6250%

Training size: 200
Error: 2.6250%

Training size: 400
Error: 1.8750%

Training size: 800
Error: 1.7500%

Training size: 1400
Error: 1.6250%