1. GDA and QDA

GDA: $\Sigma_1 = \Sigma_2$

QDA: $\Sigma_1 \neq \Sigma_2$

In GDA, we have:

$$l(\phi, \mu_0, \mu_1, \Sigma) = \log \prod_{i=1}^{m} p(x^{(i)}, y^{(i)}; \phi, \mu_0, \mu_1, \Sigma)$$

$$= \log \prod_{i=1}^{m} p(y^{(i)}; \phi) \cdot p(x^{(i)} | y^{(i)}; \mu_0, \mu_1, \Sigma)$$

$$= \sum_{i=1}^{m} \log p(x^{(i)} | y^{(i)}; \mu_0, \mu_1, \Sigma) + \sum_{i=1}^{m} \log p(y^{(i)}; \phi)$$

$$= \sum_{i=1}^{m} \left(\log \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{n}}} + \left(-\frac{1}{2}(x^{(i)} - \mu_{y^{(i)}})^T \cdot \Sigma^{-1} \cdot (x^{(i)} - \mu_{y^{(i)}})\right)\right)$$

$$+ \sum_{i=1}^{m} y^{(i)} \cdot \log \phi + (1 - y^{(i)}) \log(1-\phi)$$

Finally, we got: $\phi = \frac{1}{m} \sum_{i=1}^{m} \{y^{(i)} = 1\}$

$$\mu_b = \frac{\sum_{i=1}^{m} \{y^{(i)} = b\} \cdot x^{(i)}}{\sum_{i=1}^{m} \{y^{(i)} = b\}}$$

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

In QDA,

$$l(\phi, \mu_0, \mu_1, \Sigma_0, \Sigma_1) = \sum_{i=1}^{m} \log p(x^{(i)} | y^{(i)}; \mu_0, \mu_1, \Sigma_0, \Sigma_1) + \sum_{i=1}^{m} \log p(y^{(i)}; \phi)$$

$$= \sum_{i=1}^{m} 1\{y^{(i)}=0\}\left[\log \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma_0|^{\frac{1}{n}}} + \left(-\frac{1}{2}(x^{(i)}-u_0)^T \cdot \Sigma_0^{-1} \cdot (x^{(i)}-u_0)\right)\right]$$

$$+ \sum_{i=1}^{m} 1\{y^{(i)}=1\}\left[\log \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma_1|^{\frac{1}{n}}} + \left(-\frac{1}{2}(x^{(i)}-u_1)^T \Sigma_1^{-1}(x^{(i)}-u_1)\right)\right]$$

$$+ \sum_{i=1}^{m} \log p(y^{(i)};\phi)$$

The results are almost the same;

$$\phi = \frac{1}{m}\sum_{i=1}^{m}\{y^{(i)}=1\}$$

$$u_b = \frac{\sum_{i=1}^{m}\{y^{(i)}=b\}\cdot x^{(i)}}{\sum_{i=1}^{m}\{y^{(i)}=b\}}$$

$$\Sigma_b = \frac{\sum_{i=1}^{m}\{y^{(i)}=b\}(x^{(i)}-u_b)(x^{(i)}-u_b)^T}{\sum_{i=1}^{m}\{y^{(i)}=b\}}$$

See demo (Jupyter notebook)

2. Multi-variate Bernoulli and Multi-nomial

X	1	......	n
P	P_1		P_n

$P_1 + \cdots + P_n = 1$

$$X = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \begin{array}{l} a \\ \vdots \\ \text{casino} \\ \vdots \\ \text{I} \\ \vdots \\ \text{zyzzva} \end{array}$$

$$X = \begin{bmatrix} 0 \\ \vdots \\ 10 \\ \vdots \\ 100 \\ \vdots \\ 1 \end{bmatrix} \begin{array}{l} a \\ \vdots \\ \text{casino} \\ \\ \text{I} \\ \\ \text{zyzzyva} \end{array}$$

$n \times 1$ (n: number of features, dictionary size)

$$Len(x) = \sum_{j=1}^{n} x_j$$

$$x \mid y=b \sim M(\phi_{1|y=b} \cdots \phi_{n|y=b})$$

$$P(x \mid y=b) = \phi_{1|b}^{x_1} \cdot \phi_{2|b}^{x_2} \cdots \phi_{n|b}^{x_n}$$

$$P(x, y) = L(\phi_y, \phi_{k|y=0}, \phi_{k|y=1}) = \prod_{i=1}^{m} P(x^{(i)}, y^{(i)})$$

$$= \prod_{i=1}^{m} P(y^{(i)}) \cdot P(x^{(i)} \mid y^{(i)})$$

$$\log L = \underbrace{\sum_{i=1}^{m} \log P(y^{(i)})}_{(1)} + \underbrace{\sum_{i=1}^{m} \log P(x^{(i)} \mid y^{(i)})}_{(2)}$$

$$(2) = \sum_{i=1}^{m} \log P(x^{(i)} \mid y^{(i)})$$

$$= \sum_{i=1}^{m} \log P(x^{(i)} \mid y=1)^{y^{(i)}} \cdot P(x^{(i)} \mid y=1)^{1-y^{(i)}}$$

$$= \underbrace{\sum_{i=1}^{m} y^{(i)} \log P(x^{(i)} \mid y=1)}_{(a)} + \sum_{i=1}^{m} (1-y^{(i)}) \log P(x^{(i)} \mid y=0)$$

$$(a) = \sum_{i=1}^{m} y^{(i)} \log \prod_{j=1}^{n} \phi_{j|1}^{x_j^{(i)}}$$

$$= \sum_{i=1}^{m} y^{(i)} \sum_{j=1}^{n} x_j^{(i)} \log \phi_{j|1}$$

We have: $\phi_{j|b} \geq 0$; $\sum_{j=1}^{n} \phi_{j|b} = 1$

Lagrange multiplier.

$$L = \sum_{i=1}^{m} y^{(i)} \sum_{j=1}^{n} x_j^{(i)} \log \phi_{j|1} - \lambda \left(\sum_{j=1}^{n} \phi_{j|1} - 1 \right)$$

$$\frac{\partial L}{\ } \quad \left[\frac{\sum_{i=1}^{m} y^{(i)} x_0^{(i)}}{\ } - \lambda \right] = \left[0 \right]$$

$$\frac{\quad}{\partial \phi_{j|1}} = \begin{bmatrix} \frac{\quad}{\phi_{0|1}} - \lambda \\ \cdots\cdots \\ \frac{\sum_{i=1}^{m} y^{(i)} \cdot x_n^{(i)}}{\phi_{n|1}} - \lambda \end{bmatrix} = \begin{bmatrix} \vdots \\ 0 \end{bmatrix}$$

$$\Rightarrow \phi_{0|1} = \frac{\sum_{i=1}^{m} y^{(i)} x_0^{(i)}}{\lambda}$$

$$\because \sum_{j=1}^{n} \phi_{j|1} = \frac{\sum_{j=1}^{n} \sum_{i=1}^{m} y^{(i)} \cdot x_j^{(i)}}{\lambda} = \frac{\sum_{i=1}^{m} y^{(i)} \cdot \sum_{j=1}^{n} x_j^{(i)}}{\lambda} = \boxed{\frac{\sum_{i=1}^{m} y^{(i)} Len(x^{(i)})}{\lambda}}$$

$$= 1$$

$$\phi_{j|b} = \frac{\sum_{i=1}^{m} 1\{y^{(i)} = b\} \cdot x_j^{(i)}}{\sum_{i=1}^{m} 1\{y^{(i)} = b\} \cdot Len(x^{(i)})}$$