1. $J(\theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}\left(\sum_{j=0}^{n} x_j^{(i)}\theta_j - y^{(i)}\right)^2 + \lambda\sum_{i=1}^{n}\theta_j^2\right]$

$\because \frac{\partial}{\partial\theta_p}J = \frac{1}{2m}\left[2\sum_{i=1}^{m}\left(\sum_{j=0}^{n} x_j^{(i)}\theta_j - y^{(i)}\right)\cdot x_p^{(i)} + 2\lambda\theta_p\right] = \frac{1}{m}\left[\sum_{i=1}^{m}\left(\sum_{j=0}^{n} x_j^{(i)}\theta_j - y^{(i)}\right)\cdot x_p^{(i)} + \lambda\theta_p\cdot[p\neq 1]\right]$

$\frac{\partial}{\partial\theta_p\partial\theta_q}J = \frac{1}{m}\left[\sum_{i=1}^{m} x_p^{(i)}\cdot x_q^{(i)} + \lambda[p=q \wedge p\neq 1]\right]$

$\therefore \nabla_\theta J = \frac{1}{m} X^T(X\theta - Y) + \lambda\begin{bmatrix}0 & & \\ & 1 & \\ & & \ddots & \\ & & & 1\end{bmatrix}\theta$

$H = \frac{1}{m}\left(X^T X + \lambda\begin{bmatrix}0 & & \\ & 1 & \\ & & \ddots & \\ & & & 1\end{bmatrix}\right)$

由牛顿迭代得 $\theta_{t+1} = \theta_t - \left(X^TX + \lambda\begin{bmatrix}0 & & \\ & 1 & \\ & & \ddots & \\ & & & 1\end{bmatrix}\right)^{-1}\left[\left(X^TX + \lambda\begin{bmatrix}0 & & \\ & 1 & \\ & & \ddots & \\ & & & 1\end{bmatrix}\right)\theta_t - X^TY\right]$

$= \left(X^TX + \lambda\begin{bmatrix}0 & & \\ & 1 & \\ & & \ddots & \\ & & & 1\end{bmatrix}\right)^{-1} X^TY$

2. $l(\varphi,\mu_0,\mu_1,\Sigma) = \sum_{i=1}^{m}\log p(x^{(i)}|y^{(i)}=0)[y^{(i)}=0]$ ① $+$

$\sum_{i=1}^{m}\log p(x^{(i)}|y^{(i)}=1)[y^{(i)}=1]$ ② $+$

$\sum_{i=1}^{m}\log p(y^{(i)})$ ③

现令 $\frac{\partial}{\partial\varphi}l = 0$，只有③与 φ 有关，故有

$\frac{\partial l}{\partial\varphi} = \frac{\partial\sum_{i=1}^{m}\log p(y^{(i)})}{\partial\varphi} = \frac{\partial\sum_{i=1}^{m} y^{(i)}\cdot\log\varphi + (1-y^{(i)})\cdot\log(1-\varphi)}{\partial\varphi}$

$= \sum_{i=1}^{m}\frac{[y^{(i)}=1]}{\varphi} - \sum_{i=1}^{m}\frac{[y^{(i)}=0]}{1-\varphi} = 0$

$\varphi\left(\sum_{i=1}^{m}[y^{(i)}=1] + \sum_{i=1}^{m}[y^{(i)}=0]\right) = \sum_{i=1}^{m}[y^{(i)}=1]$

$\therefore \varphi = \frac{1}{m}\sum_{i=1}^{m}[y^{(i)}=1]$

现令 $\nabla_{\mu_0} l = 0$，只有①与 μ_0 有关，故有.

$$\nabla_{\mu_0} l = \frac{\partial l}{\partial \mu_0} = \frac{\partial \sum_{i=1}^{m} \log p(x^{(i)} | y^{(i)}=0) \cdot [y^{(i)}=0]}{\partial \mu_0}$$

$$= \frac{\partial}{\partial \mu_0} \sum_{i=1}^{m} \left[\log \frac{1}{(2\pi)^{\frac{n}{2}} \cdot |\Sigma|^{\frac{1}{2}}} - \frac{1}{2}(x^{(i)}-\mu_0)^T \Sigma^{-1} (x^{(i)}-\mu_0)\right] [y^{(i)}=0]$$

$$= \frac{\partial}{\partial \mu_0} \sum_{i=1}^{m} (x^{(i)}-\mu_0)^T \Sigma^{-1} (x^{(i)}-\mu_0) \, [y^{(i)}=0]$$

$$= \frac{\partial}{\partial \mu_0} \sum_{i=1}^{m} (-2 x^{(i)T} \Sigma^{-1} \mu_0 + \mu_0^T \Sigma^{-1} \mu_0) \, [y^{(i)}=0]$$

$$= \sum_{i=1}^{m} (-2 x^{(i)} \Sigma^{-1} + 2\mu_0^T \Sigma^{-1}) \cdot [y^{(i)}=0] = 0$$

$$\therefore \quad \mu_0^T \cdot \sum_{i=1}^{m} [y^{(i)}=0] = \sum_{i=1}^{m} x^{(i)T} [y^{(i)}=0]$$

$$\mu_0 = \frac{\sum_{i=1}^{m} x^{(i)} [y^{(i)}=0]}{\sum_{i=1}^{m} [y^{(i)}=0]}$$

同理 $$\mu_1 = \frac{\sum_{i=1}^{m} x^{(i)} [y^{(i)}=1]}{\sum_{i=1}^{m} [y^{(i)}=1]}$$

现令 $\nabla_\Sigma l = 0$，只有①②与 Σ 有关，故.

$$\frac{\partial l}{\partial \Sigma} = \frac{\partial}{\partial \Sigma} \sum_{i=1}^{m} \log p(x^{(i)} | y^{(i)})$$

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

$$= \frac{\partial}{\partial \Sigma} \sum_{i=1}^{m} -\frac{1}{2}\log|\Sigma| - \frac{1}{2}(x^{(i)}-\mu_{y^{(i)}})^T \Sigma^{-1} (x^{(i)}-\mu_{y^{(i)}})$$

$$= \sum_{i=1}^{m} -\frac{1}{2} \frac{1}{|\Sigma|} \cdot |\Sigma| \cdot (\Sigma^{-1})^T + \frac{1}{2} \Sigma^{-1} (x^{(i)}-\mu_{y^{(i)}})(x^{(i)}-\mu_{y^{(i)}})^T \Sigma^{-1} = 0.$$

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

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

综上：
$$\varphi = \frac{1}{m}\sum_{i=1}^{m}[y^{(i)}=1]$$
$$\mu_0 = \sum_{i=1}^{m}x^{(i)}[y^{(i)}=0] \Big/ \sum_{i=1}^{m}[y^{(i)}=0]$$
$$\mu_1 = \sum_{i=1}^{m}x^{(i)}[y^{(i)}=1] \Big/ \sum_{i=1}^{m}[y^{(i)}=1]$$
$$\Sigma = \frac{1}{m}\sum_{i=1}^{m}(x^{(i)}-\mu_{y^{(i)}})(x^{(i)}-\mu_{y^{(i)}})^T$$

3. (i) 目标： $\max \sum_{y\in Y} C_y \log P_y$ s.t. $P_y \geq 0,\ \forall y\in Y$
$\sum_{y\in Y} P_y = 1$

构造函数 $f(\lambda, p) = \sum_{y\in Y} C_y \log P_y + \lambda\left(\sum_{y\in Y} P_y - 1\right)$

$\therefore$ 有 $\begin{cases} \frac{\partial}{\partial P_y} f = 0, & \forall y\in Y \\ \sum_{y\in Y} P_y = 1 \end{cases}$

$\because \frac{\partial}{\partial P_y} f = \frac{C_y}{P_y} + \lambda \Rightarrow P_y = -\frac{C_y}{\lambda}$

$\Rightarrow \lambda = -\sum_{y\in Y} C_y$

$\therefore P_y = \frac{C_y}{\sum_{y\in Y} C_y} = \frac{C_y}{N}$

(ii) 对NB有对数似然函数 $l = \sum_{i=1}^{m}\log P(y^{(i)}) + \sum_{i=1}^{m}\sum_{j=0}^{n}\log P_j(x_j^{(i)}|y^{(i)})$

令 $l_1 = \sum_{i=1}^{m}\log P(y^{(i)})$, $l_2 = \sum_{i=1}^{m}\sum_{j=0}^{n}\log P_j(x_j^{(i)}|y^{(i)})$

$\therefore l = l_1 + l_2$，其中 l_1 只与 $p(y)$ 有关，l_2 只与 $P_j(x|y)$ 有关.

$\therefore p(y) = \arg\max_{P_y} \sum_{i=1}^{m}\log p(y^{(i)})$ 其中 $C_i = 1\ \forall i = 1,2,\dots,m.$

$\therefore p^*(y) = \frac{\sum_{i=1}^{m} C_i [y^{(i)}=y]}{N} = \frac{\sum_{i=1}^{m}[y^{(i)}=y]}{m}.$

而 $\forall j=0,1,2,\cdots\cdots,n$； $i=1,2,3,\cdots,m$：

$$P_j^*(x|y)=\underset{P_j(x|y)}{\arg\max}\sum_{i=1}^{m}[y^{(i)}=y]\cdot\log P_j(x_j^{(i)}|y)$$ 其中 $C_i=[y^{(i)}=y]$

由上式知，$$P_j^*(x|y)=\frac{\sum_{i=1}^{m}[y^{(i)}=y]\cdot[x_j^{(i)}=x]}{\sum_{i=1}^{m}[y^{(i)}=y]}$$