Answer:

intro

We seek to maximise $like(\mu_0, \mu_1, \Sigma, \phi) = p(x, y|\mu_0, \mu_1, \Sigma, \phi) = \Pi p(y_i|\phi)p(x_i|\mu_0, \mu_1, \Sigma, y_i)$, using simple formula $p(x_i, y_i) = p(y_i)p(x_i|y_i)$, and that $p(y|\phi, \mu_0, ...) = p(y|\phi)$. Then we can separately maximise $\Pi p(y_i|\phi)$ and $\Pi p(x_i|\mu_0, \mu_1, \Sigma, y_i))$, denoting each (not sure if this is standard) $like(\phi), like(\mu_0, \mu_1, \Sigma)$.

MLE of ϕ

$like(\phi) = p(y|\phi) = \Pi_{i=1}^n \phi^{y_i}(1-\phi)^{1-y_i}$, where $y_i = 0, 1$. Then to maximise log likelihood, $l(\phi) = \sum y_i \log(\phi) + (1 - y_i)\log(1-\phi)$ we seek $\frac{dl(\phi)}{d\phi} = 0$, i.e. $\sum \frac{y_i}{\phi} + \frac{y_i - 1}{1-\phi} = 0 = \sum \frac{y_i - \phi}{\phi(1-\phi)}$ which coincides with $\phi = \bar{y}$.

MLE of μ_0, μ_1

$like(\mu_0, \mu_1, \Sigma) = \Pi p(x_i|\mu_0, \mu_1, \Sigma, y_i))$, so we'll maximise over log likelihood i.e. over

$$-\frac{nd}{2}\log(2\pi) - \frac{n}{2}\log(|\Sigma|) - \frac{1}{2}\Sigma_{y_i=0}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0) - \frac{1}{2}\Sigma_{y_i=1}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)$$

Then the derivative of the log likelihood with respect to μ_0 is $\Sigma_{y_i=0}\Sigma^{-1}(x-\mu_0)$ i.e. is maximised with $\mu_0 = \bar{x}$ averaged over $x_i : y_i = 0$, and similarly for μ_1.

MLE of Σ

To maximise over Σ seems pretty involved. We'll require a formula for $\frac{\partial}{\partial A_{ij}}\det(A)$. We'll use the matrix minors formulas: $\delta_{ik}|A| = \sum_{j=1}^n |A_{ji}|A_{jk}$ Where A_{ji} is the minor matrix, and $|A_{ji}|$ is its determinant. This gives a form of the inverse for A of $K_{ij} = \frac{(-1)^{i+j}}{|A|}|A_{ji}|$ as this gives $K_{ij}A_{jk} = \sum_{j=1}^n \frac{(-1)^{i+j}}{|A|}|A_{ji}|A_{jk} = \delta_{ij}$. Then $\frac{\partial}{\partial A_{ab}}|A| = \frac{\partial}{\partial A_{ab}}\left(\sum_{j=1}^n(-1)^{b+j}|A_{jb}|A_{jb}\right)$. The sum only has one term with a A_{ab} component, where $j = a$, so we get $\frac{\partial}{\partial A_{ab}}|A| = \frac{\partial}{\partial A_{ab}}(-1)^{a+b}|A_{ab}|A_{ab} = (-1)^{a+b}|A_{ab}| = A^{-1}_{ba}|A|$ (by using the minor form of matrix inverse). This gives us the formula $\nabla_A|A| = |A|(A^{-1})^T$.
So seeking a solution of

$$\nabla_A\left(\frac{n}{2}\log(|A|) - \frac{1}{2}\Sigma_{y_i=0}(x-\mu_0)^T A(x-\mu_0) - \frac{1}{2}\Sigma_{y_i=1}(x-\mu_1)^T A(x-\mu_1)\right) = 0$$

where we let $A = \Sigma^{-1}$ and use $|\Sigma| = 1/|A|$, as well as assuming μ_0, μ_1 are already maximised (i.e. $\mu_0 = \hat{\mu_0}_{MLE}$ etc.). Then we get $\frac{n}{2}A^{-1} - \frac{1}{2}\Sigma_{i=1}^n(x-\mu_{y_i})(x-\mu_{y_i})^T = 0$ which gives the desired MLE equation, where we used that Σ, and so A too, are symmetric so $(A^{-1})^T = A^{-1}$, and that for any matrix A, $\frac{\partial}{\partial A}z^T Az = zz^T$.