The task is to proove that the solutions to maximize the log likelihood for GDA are those given in the lectures. For the sake of clarity we will make some definitions. Let's note the the number of samples corresponding to class one $y^{(i)} = 1$ as k, and class zero $y^{(i)} = 1$ as $m - k$, while total number of samples are m. Let's also define a subset of indices for class 1 as $C_1 = \{i;\ y^{(i)} = 1\}$, and for class 0 $C_0 = \{i;\ y^((i)) = 0\}$. Also for this task we may consider that the data is one-dimensional, so the covariance matrix $\Sigma = \sigma^2$ is just a real number. So the equations we need to proove are listed below.

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

$$\mu_0 = \frac{\sum\limits_{i \in C_0} x^{(i)}}{k}$$

$$\mu_1 = \frac{\sum\limits_{i \in C_1} x^{(i)}}{m - k}$$

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

$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m}(x^{(i)} - \mu_{y^{(i)}})^2$$

Lets first start by simplifiying by reminding the conditional probability distribution functions for each class:

$$\begin{aligned}
p(x|y^{(i)} = 0) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\Big[-\frac{1}{2}\frac{(x-\mu_0)^2}{\sigma^2}\Big] = p_0(x) \\
p(x|y^{(i)} = 1) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\Big[-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma^2}\Big] = p_1(x)
\end{aligned}$$

The log-likelihood can be written as:

$$\begin{aligned}
&\ell(\phi, \mu_0, \mu_1, \sigma) = \\
log\mathcal{L}(\phi, \mu_0, \mu_1, \sigma) &= log\prod_{i=1}^{m} p(x^{(i)}, y^{(i)}; \phi, \mu_0, \mu_1, \sigma) = \sum_{i=1}^{m} log\ p(x^{(i)}|y^{(i)}; \mu_0, \mu_1, \sigma)\ p(y^{(i)}; \phi) \\
&= \sum_{i \in C_0} log(p_0(x^{(i)})(1-\phi)) + \sum_{i \in C_1} log(p_1(x^{(i)})\phi) \\
&= \sum_{i \in C_0} logp_0(x^{(i)}) + \sum_{i \in C_1} logp_1(x^{(i)}) + (m-k)log(1-\phi) + k \cdot log\phi
\end{aligned}$$

To find a maximum of log-likelihood w.r.t. ϕ we can just differentiate and find a solution to eqation:

$$\frac{d\ell}{d\phi} = 0$$

$$\begin{aligned}
\frac{d}{d\phi}\Big[(m-k)log(1-\phi) + k \cdot log\phi\Big] &= 0 \\
k\frac{1}{\phi} - (m-k)\frac{1}{1-\phi} &= 0 \\
1 - (\frac{m}{k} - 1) \cdot \frac{\phi}{1-\phi} &= 0 \\
\frac{\phi}{(1-\phi)} &= \frac{k}{m-k} \\
\frac{\phi}{(1-\phi)} &= \frac{k/m}{1-k/m} \\
\phi &= \frac{k}{m}
\end{aligned}$$

Now let's find analogously derivative of log-likelihood w.r.t to other parameters.

$$logp_1(x^{(i)}) = \frac{1}{const} - \frac{1}{2\sigma^2}(x-\mu_1)^2$$

$$\frac{d\ell}{d\mu_1} = 0$$

$$\frac{d\ell}{d\mu_1} = \sum_{i \in C_1} \frac{\partial logp_1(x^{(i)})}{\partial \mu_1} = \frac{1}{2\sigma^2} \sum_{i \in C_1} \frac{\partial (x^{(i)} - \mu_1)^2}{\partial \mu_1} =$$
$$= \sum_{i \in C_1} \frac{d(x^2 - 2x\mu_1 + \mu_1^2)}{d\mu_1} = \sum_{i \in C_1} -2x + 2\mu_1 = 0$$

$$\sum_{i \in C_1} x^{(i)} = \sum_{i \in C_1} \mu_1$$

$$\sum_{i \in C_1} x^{(i)} = k\mu_1$$

$$\mu_1 = \frac{\sum_{i \in C_1} x^{(i)}}{k}$$

And the last one σ. First lets simplify the log-likelihood equation containing σ:

$$\sum_{i \in C_1} logp_1(x^{(i)}) + \sum_{i \in C_0} logp_0(x^{(i)}) =$$

$$= \sum log \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2} \sum_{i \in C_1} \frac{(x-\mu_1)^2}{\sigma^2} - \frac{1}{2} \sum_{i \in C_0} \frac{(x-\mu_0)^2}{\sigma^2} \tag{1}$$

$$\sum log \frac{1}{\sqrt{2\pi}\sigma} = -\sum log\sqrt{2\pi} - \sum log\sigma = -m \cdot log\sqrt{2\pi} - m \cdot log\sigma \tag{2}$$

Insert (2) into (1) and differentiating w.r.t σ woudl give us

$$\frac{d\ell}{d\sigma} = -m\frac{dlog\sigma}{\sigma} - -\frac{1}{\cancel{2}} \sum_{i \in C_1} \frac{(x-\mu_1)^2}{\sigma^3} \cdot (-\cancel{2}) - \frac{1}{\cancel{2}} \sum_{i \in C_1} \frac{(x-\mu_1)^2}{\sigma^3} \cdot (-\cancel{2})$$

$$\frac{d\ell}{d\sigma} = -m\frac{1}{\cancel{\sigma}} + \sum_{i \in C_1} \frac{(x-\mu_1)^2}{\sigma^{\cancel{3}}} + \sum_{i \in C_0} \frac{(x-\mu_0)^2}{\sigma^{\cancel{3}}} = 0$$

$$\frac{\sum_{i \in C_1}(x-\mu_1)^2 + \sum_{i \in C_1}(x-\mu_0)^2}{\sigma^2} = m$$

$$\Sigma = \sigma^2 = \frac{1}{m}\sum_{i=1}^{m}(x-\mu_{y^{(i)}})^2$$