1.15

$Z = \{0, 1\}$ where 0 indicates that the student didn't understand the topic and 1 that they did.
The probability function p_Z tells us how likely it is for any given student to have understood the topic or not without further any information on the student.
$p_Z(0) = p_0$ $p_Z(1) = p_1$
Since the points are Gaussian distributed we get $X = \mathbb{R}$.
$p_{X|Z}$ is the probability of any given student to score a certain number of points given they either understood the topic or not.
Since the points are Gaussian distributed we also get the probability of any x
inX simply by calculating the formula for a Gaussian at point x with the mean depending on whether the student has understood the topic or not.

$$
\begin{aligned}
p_{X|Z}(x \in X|0) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu_0)^2}{2\sigma^2}\right) \\
p_{X|Z}(x \in X|1) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma^2}\right)
\end{aligned}
$$

Where μ_0 is the mean for students who have not understood the topic and μ_1 for those who did.

1.16

$$
\begin{aligned}
p_{Z|X}(1|x) &= p_{X|Z}(x|1) \cdot \frac{p_Z(1)}{p_X(x)} \\
&= \frac{p_{X|Z}(x|1) \cdot p_Z(1)}{p_{X|Z}(x|0) \cdot p_Z(0) + p_{X|Z}(x|1) \cdot p_Z(1)} \\
&= \frac{\exp\left(-\frac{(x-\mu_1)^2}{2\sigma^2}\right) \cdot p_1}{\exp\left(-\frac{(x-\mu_0)^2}{2\sigma^2}\right) \cdot p_0 + \exp\left(-\frac{(x-\mu_1)^2}{2\sigma^2}\right) p_1} \\
&= \frac{1}{1 + \exp\left(\frac{(x-\mu_1)^2-(x-\mu_0)^2}{2\sigma^2}\right) \cdot \frac{p_0}{p_1}} \\
&= \frac{1}{1 + \exp\left(\frac{2x\mu_0-2x\mu_1+\mu_1^2-\mu_0^2}{2\sigma^2}\right) \cdot \frac{p_0}{p_1}}
\end{aligned}
$$

1.17

$$
\begin{aligned}
-\ln(x_i) &= -\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)\right)\\
&= -\ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \ln\left(\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)\right)\\
&= -\ln(1) + \ln(\sqrt{2\pi}\sigma) + \frac{(x_i-\mu)^2}{2\sigma^2}\\
&= \ln(\sqrt{2\pi}\sigma) + \frac{(x_i-\mu)^2}{2\sigma^2}\\
-\sum_{i=1}^{m}\ln(x_i) &= \sum_{i=1}^{m}\left(\ln(\sqrt{2\pi}\sigma) + \frac{(x_i-\mu)^2}{2\sigma^2}\right)\\
&= m\cdot\ln(\sqrt{2\pi}\sigma) + \frac{1}{\sigma^2}\sum_{i=1}^{m}\frac{(x_i-\mu)^2}{2}
\end{aligned}
$$

To optimize in relation tu μ one must derivative in relation to μ and then equal that to zero. We don't need to check the second derivative since we can assume ln() to be convex.

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\mu}\left(-\sum_{i=1}^{m}\ln(x_i)\right) &= 0\\
\Leftrightarrow \frac{\mathrm{d}}{\mathrm{d}\mu}\left(m\cdot\ln(\sqrt{2\pi}\sigma) + \frac{1}{\sigma^2}\sum_{i=1}^{m}\frac{(x_i-\mu)^2}{2}\right) &= 0\\
\Leftrightarrow \frac{1}{\sigma^2}\sum_{i=1}^{m}\mu - x_i &= 0\\
\Leftrightarrow m\mu &= \sum_{i=1}^{m}x_i\\
\Leftrightarrow \mu &= \frac{1}{m}\sum_{i=1}^{m}x_i
\end{aligned}
$$

Same for σ

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\sigma}\left(-\sum_{i=1}^{m}\ln(x_i)\right) &= 0\\
\Leftrightarrow \frac{\mathrm{d}}{\mathrm{d}\sigma}\left(m\cdot(\ln(\sqrt{2\pi}) + \ln(\sigma)) + \frac{1}{\sigma^2}\sum_{i=1}^{m}\frac{(x_i-\mu)^2}{2}\right) &= 0\\
\Leftrightarrow \frac{m}{\sigma} - \frac{2}{\sigma^3}\sum_{i=1}^{m}\frac{(x_i-\mu)^2}{2} &= 0\\
\Leftrightarrow \frac{m}{\sigma} &= \frac{1}{\sigma^3}\sum_{i=1}^{m}(x_i-\mu)^2\\
\sigma^2 &= \frac{1}{m}\sum_{i=1}^{m}(x_i-\mu)^2
\end{aligned}
$$