(a)

The log-likelihood is

$$\ell(\theta) \;=\; \log L(\theta) \;=\; \sum_{i=1}^{m}\Big[y^{(i)}\log h\big(x^{(i)}\big) + \big(1-y^{(i)}\big)\log\big(1-h\big(x^{(i)}\big)\big)\Big].$$

After training, the gradients are zero:

$$\frac{\partial \ell(\theta)}{\partial \theta_j} \;=\; \sum_{i=1}^{m}\big(y^{(i)} - h\big(x^{(i)}\big)\big)\,x_j^{(i)} \;=\; 0.$$

Set $j=0$. Since $x_0^{(i)} = 1$ for all i, we get

$$\sum_{i=1}^{m}\big(y^{(i)} - h\big(x^{(i)}\big)\big) \;=\; 0 \;\Longrightarrow\; \sum_{i=1}^{m} h\big(x^{(i)}\big) \;=\; \sum_{i=1}^{m} y^{(i)}.$$

But

$$h\big(x^{(i)}\big) = P\big(y^{(i)} = 1 \mid x^{(i)};\theta\big), \quad y^{(i)} = \mathbf{1}\{y^{(i)} = 1\},$$

so

$$\sum_{i=1}^{m} P\big(y^{(i)} = 1 \mid x^{(i)};\theta\big) \;=\; \sum_{i=1}^{m}\mathbf{1}\{y^{(i)} = 1\}.$$

Finally, when $(a,b) = (0,1)$, one has $I_{a,b} = \{1,\dots,m\}$ and $|I_{a,b}| = m$, giving

$$\frac{1}{m}\sum_{i\in I_{a,b}} P\big(y^{(i)} = 1 \mid x^{(i)};\theta\big) \;=\; \frac{1}{m}\sum_{i\in I_{a,b}}\mathbf{1}\{y^{(i)} = 1\},$$

as required.

(b)

The model is perfectly calibrated doesn't necessarily imply that the model achieves perfect accuracy.

The converse is also not necessarily true.

Assume that $(a,b) = (0.5, 1)$.

When the model achieves perfect accuracy, the predictions are all correct, i.e.

$$\sum_{i\in I_{a,b}} \mathbb{I}\{y^{(i)} = 1\} = |\{i \in I_{a,b}\}|$$

For all $i \in I_{a,b}$

$$0.5 < P(y^{(i)} = 1 \mid x^{(i)}; \theta) < 1$$

So

$$\frac{\sum_{i \in I_{a,b}} P(y^{(i)} = 1 \mid x^{(i)}; \theta)}{|\{i \in I_{a,b}\}|} < \frac{\sum_{i \in I_{a,b}} \mathbb{I}\{y^{(i)} = 1\}}{|\{i \in I_{a,b}\}|}$$

However, when the model is perfectly calibrated, the following property always holds

$$\frac{\sum_{i \in I_{a,b}} P(y^{(i)} = 1 \mid x^{(i)}; \theta)}{|\{i \in I_{a,b}\}|} = \frac{\sum_{i \in I_{a,b}} \mathbb{I}\{y^{(i)} = 1\}}{|\{i \in I_{a,b}\}|}$$

So model is perfectly calibrated doesn't mean model achieves perfect accuracy. The converse neither.

(c)

Apply ℓ_2-regularization to the cost function:

$$J(\theta) = -\sum_{i=1}^{m} \Big[y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log\big(1 - h_\theta(x^{(i)})\big) \Big] \; + \; \frac{\lambda}{2} \|\theta\|_2^2.$$

Then the gradient condition becomes

$$\frac{\partial J(\theta)}{\partial \theta_j} = -\sum_{i=1}^{m} \big(y^{(i)} - h_\theta(x^{(i)})\big)\, x_j^{(i)} \; + \; \lambda\, \theta_j = 0.$$

Obviously, if $\theta_0 = 0$ the calibration property still holds; otherwise, it does not.