Solution 1:

(a) The logistic function is a special case of the softmax for two classes. We have

$$\pi_1(x) = \frac{\exp(\theta_1^\top x)}{\exp(\theta_1^\top x) + \exp(\theta_2^\top x)}$$

and

$$\pi_2(x) = \frac{\exp(\theta_2^\top x)}{\exp(\theta_1^\top x) + \exp(\theta_2^\top x)}.$$

We get:

$$\pi_1(x) = \frac{1}{(\exp(\theta_1^\top x) + \exp(\theta_2^\top x)) / \exp(\theta_1^\top x)} = \frac{1}{\exp((\theta_1 - \theta_1)^\top x) + \exp((\theta_2 - \theta_1)^\top x)} = \frac{1}{1 + \exp(\theta^\top x)}$$

where $\theta = \theta_2 - \theta_1$ and $\pi_2(x) = 1 - \pi_1(x)$.

(b) For g classes and $n = 1$ trials (actually we are dealing with a multinoulli or categorial distribution), the likelihood $l(\boldsymbol{\pi})$ of a single observation y is given by

$$l(\boldsymbol{\pi}) = \prod_{k=1}^{g} \pi_k^{\mathbb{1}_{\{y=k\}}}.$$

Now let's look at the logarithmic loss in softmax regression:

$$\text{MC logloss} = -\sum_{k=1}^{g} \mathbb{1}_{\{y=k\}} \log \pi_k.$$

This is in fact just the negative logarithm of our likelihood: $-\log l(\boldsymbol{\pi}) = -\sum_{k=1}^{g} \mathbb{1}_{\{y=k\}} \log \pi_k$.