Solution 1:

1) In the case of the linear model (LM), empirical risk minimization (ERM) does not necessarily result in a trained model that always satisfies $\hat{\boldsymbol{\theta}}^T\mathbf{x} \in [0,1]$, thus leading to predictions that cannot be sensibly interpreted as probabilities. Therefore, the hypothesis space must be restricted to a function that ensures above condition, which holds for the logistic function s:

$$\mathcal{H} = \left\{\pi : \mathcal{X} \to [0,1] \mid \pi(\mathbf{x}) = s(\boldsymbol{\theta}^\top \mathbf{x})\right\} \tag{1}$$

2) If one plugs in the Bernoulli loss function $L\left(y, \pi(\mathbf{x})\right)$ into the empirical risk function $\mathcal{R}_{\text{emp}}(f)$, lets probabilities $\pi(\mathbf{x})$ be modeled by the logistic function $\pi(\mathbf{x} \mid \boldsymbol{\theta}) = s(\boldsymbol{\theta}^\top \mathbf{x})$, and specifies the risk surface to be minimized with regards to the parameter vector $\boldsymbol{\theta} \in \Theta$, the following explicit ERM problem emerges:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \Theta} \sum_{i=1}^{n} - y^{(i)} \ln\left(s\left(\boldsymbol{\theta}^T \mathbf{x}^{(i)}\right)\right) - \left(1 - y^{(i)}\right) \ln\left(1 - s\left(\boldsymbol{\theta}^T \mathbf{x}^{(i)}\right)\right) \tag{2}$$

4) Deriving the log-likelihood function ℓ of a single Bernoulli distributed random variable Y, one gets

$$\begin{aligned} \mathcal{L} &= \mathbb{P}(Y = y) = \pi^y (1 - \pi)^{1-y} && (3) \\ \ell &= ln(\mathcal{L}) && (4) \\ &= y \ln(\pi) + (1 - y) \ln(1 - \pi), && (5) \end{aligned}$$

which is equivalent to the Bernoulli loss function if one multiplies by (-1). This demonstrates the correspondence of *maximum* likelihood estimation and empirical risk *minimization* in the context of a logistic regression model. Both approaches lead to identical parameter estimates.