Large Scale Machine Learning and Deep Learning
Review Questions 2

1. What is cross validation and how does it work?

   Answer: the regression slides, pages 73–75

---

2. Mathematically show that the softmax function with two classes ($\mathtt{k} = 2$) is equivalent to the sigmoid function?
   Answer: let's look at the multiclass logistic regression, with $k = 2$:

   $$\mathtt{p}(\mathtt{y}^{(\mathtt{i})} = 0) = \frac{\mathtt{e}^{\mathbf{w}_0^T \mathbf{x}^{(\mathtt{i})}}}{\mathtt{e}^{\mathbf{w}_0^T \mathbf{x}^{(\mathtt{i})}} + \mathtt{e}^{\mathbf{w}_1^T \mathbf{x}^{(\mathtt{i})}}} = \frac{\mathtt{e}^{(\mathbf{w}_0 - \mathbf{w}_1)^T \mathbf{x}^{(\mathtt{i})}}}{1 + \mathtt{e}^{(\mathbf{w}_0 - \mathbf{w}_1)^T \mathbf{x}^{(\mathtt{i})}}} = \frac{\mathtt{e}^{-\mathbf{w}^T \mathbf{x}^{(\mathtt{i})}}}{1 + \mathtt{e}^{-\mathbf{w}^T \mathbf{x}^{(\mathtt{i})}}}$$

   where $\mathbf{w} = -(\mathbf{w}_0 - \mathbf{w}_1)$, then we have:

   $$\mathtt{p}(\mathtt{y}^{(\mathtt{i})} = 1) = 1 - \mathtt{p}(\mathtt{y}^{(\mathtt{i})} = 0) = 1 - \frac{\mathtt{e}^{-\mathbf{w}^T \mathbf{x}^{(\mathtt{i})}}}{1 + \mathtt{e}^{-\mathbf{w}^T \mathbf{x}^{(\mathtt{i})}}} = \frac{1}{1 + \mathtt{e}^{-\mathbf{w}^T \mathbf{x}^{(\mathtt{i})}}}$$

---

3. As you know, in binomial logistic regression the `cost` between the true value $\mathtt{y}$ and the predicted value $\hat{\mathtt{y}}$ is measured as below:

   $$\mathtt{cost}(\hat{\mathtt{y}}, \mathtt{y}) = \begin{cases} -\mathtt{log}(\hat{\mathtt{y}}) & \text{if} \quad \mathtt{y} = 1 \\ -\mathtt{log}(1 - \hat{\mathtt{y}}) & \text{if} \quad \mathtt{y} = 0 \end{cases}$$

   Explain why $-\mathtt{log}$ is a proper function to compute the cost in logistic regression?

   Answer: the classification slides, pages 17–18

---

4. How are logistic regression cost, cross-entropy, and negative log-likelihood related?

   Answer: the classification slides, pages 30–36

---

5. Explain how a ROC curve works?

   Answer: the classification slides, pages 63–64