Name: ______________________ Roll Number: ______________________

1. (5 points) In the binary cross-entropy function

$$J(\theta) = -\frac{1}{N}\sum_{n=1}^{N} y_n \ln(\hat{y}_n) + (1 - y_n)\ln(1 - \hat{y}_n) \quad (1)$$

   (a) θ is <u>all the learnable parameters</u>.
   (b) N is <u>the number of training examples</u>.
   (c) y_n is <u>the actual class of training sample $\mathbf{x}_n$</u>.
   (d) $\hat{y}_n$ is <u>the probability of class 1 given input sample $\mathbf{x}_n$</u>.
   (e) the expression inside the sum selects the -ve log probability of <u>the correct class</u>.

2. (5 points) (a) Describe the output layer of neural networks for the following problems. Your description must include i) the number of neurons, and ii) the type of activation functions.

   i. (1 point) Classification of images of chairs, sofas and tables.

   3 neurons with softmax activations.

   ii. (1 point) Learning a vector function $\mathbf{f} \in \mathbb{R}^{13}$.

   13 neurons with linear activations.

   (b) (3 points) The softmax function for K inputs $a_1, a_2, \ldots, a_K$ is written as

$$y_k = \frac{e^{a_k}}{\sum_{j=1}^{K} e^{a_j}}$$

   Prove that the softmax function outputs multiclass probabilities. You must show that

   1. each output $y_k \geq 0$,
   2. each output $y_k \leq 1$, and
   3. sum of outputs $y_1, \ldots, y_K$ is exactly 1.

   1. $y_k > 0$ since $e^x > 0 \;\forall\, x$ and therefore numerator and denominator are both positive.
   2. $y_k \leq 1$ since denominator contains the numerator plus some non-negative terms.
   3. Sum of all outputs is

$$\begin{aligned} & y_1 + y_2 + \cdots + y_K \\ &= \frac{e^{a_1}}{\sum_{j=1}^{K} e^{a_j}} + \frac{e^{a_2}}{\sum_{j=1}^{K} e^{a_j}} + \cdots + \frac{e^{a_K}}{\sum_{j=1}^{K} e^{a_j}} \\ &= \frac{\sum_{k=1}^{K} e^{a_k}}{\sum_{j=1}^{K} e^{a_j}} \\ &= 1 \end{aligned}$$

   Since $0 \leq y_k \leq 1 \;\forall\, k$ and since $\sum_{k=1}^{K} y_k = 1$, outputs of the softmax function represent multiclass probabilities.