Work out the following problems and show all of your work.

1) Suppose we have the perceptron model below:

\[ \phi(w^T x) = \begin{cases} 
1 & \text{if } w^T x \geq 0 \\
0 & \text{if } w^T x < 0 
\end{cases} \]

\[ w^T x = 1 + x_1 - 2x_2. \]

a. What value will the perceptron model predict when \( x_1 = 9, x_2 = 5.1 \)? (5 points)

b. With the perceptron model, update the weights when we have the training observation: \( y = 1, x_1 = 9, x_2 = 5.1 \) with \( \eta = 0.1 \) and the perceptron learning rule: (5 points)

\[ w_j := w_j + \Delta w_j \]

\[ \Delta w_j = \eta (y^{(i)} - \hat{y}^{(i)}) x_j^{(i)}. \]

2) Suppose we have the sigmoid model below:

\[ \phi(w^T x) = \frac{1}{1 + e^{-w^T x}} \quad \hat{y} = \begin{cases} 
1 & \text{if } \phi(w^T x) \geq 0.5 \\
0 & \text{if } \phi(w^T x) < 0.5 
\end{cases} \]

\[ w^T x = 1 + x_1 - 2x_2. \]

a. What value will the sigmoid model predict when \( x_1 = 9, x_2 = 5.1 \)? (5 points)

b. With the sigmoid model, update the weights when we have the training observation: \( y = 1, x_1 = 9, x_2 = 5.1 \) with \( \eta = 0.1 \) and the learning rule: (5 points)

\[ w_j := w_j + \Delta w_j \]

\[ \Delta w_j = \eta (y^{(i)} - \phi(w^T x^{(i)})) x_j^{(i)}. \]
3) Using Figure 1, answer the following questions. Assume there are no bias terms contributing to the input nor any of the activations in each layer.
   a. How many layers (L) are there in this neural network? (2 point)
   
   b. How many classes will this neural network classify? (2 point)
   
   c. What are the dimensions of $W^{(2)}$? (4 points)
   
   d. State the equation for $a^{(3)}_2$. (4 points)

4) Use Figure 2 as a model to diagnose diabetic retinopathy by answering the questions below. The estimate of "unhealthy" is denoted by class 0 while "healthy" is denoted by class 1. Given the observation $x_1 = 1.08$ and $x_2 = 10$, find the following:
   a. the estimated probability that a retina is healthy, (6 point)
   
   b. the estimated probability that a retina is unhealthy, and (3 point)
   
   c. the class we would expect the observation to belong. (1 point)
5) Use Figure 2 as a model to diagnose diabetic retinopathy by answering the questions below. The estimate of “unhealthy” is denoted by class 0 while “healthy” is denoted by class 1. Given the observation $x_1 = 1.02$ and $x_2 = 3$, find the following:
   a. the estimated probability that a retina is healthy, (6 points)
   b. the estimated probability that a retina is unhealthy, and (3 points)
   c. the class we would expect the observation to belong. (1 point)

6) Use neural networks to create the following logical operators.
   a. Create a “NAND” operator using a single neuron such that we get the following output:
      (7 points)
      $$
      \begin{array}{c|c|c}
      x_1 & x_2 & y_1 \\
      \hline
      0 & 0 & 1 \\
      0 & 1 & 1 \\
      1 & 0 & 1 \\
      1 & 1 & 0 \\
      \end{array}
      $$
   b. Create an “OR” operator using a single neuron such that we get the following output:
      (7 points)
      $$
      \begin{array}{c|c|c}
      x_1 & x_2 & y_2 \\
      \hline
      0 & 0 & 0 \\
      0 & 1 & 1 \\
      1 & 0 & 1 \\
      1 & 1 & 0 \\
      \end{array}
      $$
   c. Create an “exclusive-OR” operator using the neurons from both a and b with an “AND” operator such that we get the following output: (9 points)
      $$
      \begin{array}{c|c|c}
      x_1 & x_2 & y_3 \\
      \hline
      0 & 0 & 0 \\
      0 & 1 & 1 \\
      1 & 0 & 1 \\
      1 & 1 & 0 \\
      \end{array}
      $$
7) For this problem, we will be working with the MNIST data set and making submissions to Kaggle. You can view the competition and download the data from “kaggle” at https://www.kaggle.com/c/digit-recognizer. For the competition, you will be applying logistic regression (sigmoid neurons) for different values of $C$ to the data set and submitting your predictions to kaggle. Also, you should normalize the data using the standard scaler before performing logistic regression on the data set. For each problem, hand in your code (10 points total) and specify the classification accuracy (5 points each).
   a. For $C = 10000$:
      classification accuracy: _____________
   b. For $C = 100$:
      classification accuracy: _____________
   c. For $C = 1$:
      classification accuracy: _____________