1. Recall that the perceptron algorithm cycles through the training examples, applying the following rule:

$$
\begin{aligned}
z^{(i)} &\leftarrow w^T x^{(i)} && (1)\\
\text{If} \quad & z^{(i)} t^{(i)} \leq 0, && (2)\\
& w \leftarrow w + t^{(i)} x^{(i)} && (3)
\end{aligned}
$$

(Recall that the targets take values in $\{-1, 1\}$ Suppose we make the inequality strict in the conditional, i.e. we update the weights only if $z^{(i)} t^{(i)} < 0$. What would go wrong?

You may assume the weights are initialized to 0. Hint: what happens on the first training example?

Solution: For the first training example, z evaluates to 0, so the strict version of the inequality will not be triggered, and then the weights won't be updated. Then the weights will still be zero for the second example, and so on, so the weights never actually get updated.

2. You want to train the following model using gradient descent. Here, the input x and target t are both scalar-valued.

$$
\begin{aligned}
z &= w_0 + w_1 x + w_2 x_2 && (4)\\
y &= 1 + e^z && (5)\\
\mathcal{L} &= \frac{1}{2}(log\ y - log\ t)^2 : && (6)
\end{aligned}
$$

Determine the backprop rules which will let you compute the loss derivative $\partial \mathcal{L} / \partial w_2$. Your equations should refer to previously computed values (e.g. your formula for $\bar{z}$ should be a function of $\bar{y}$). You do not need to show your work, but it may help you get partial credit. The dummy step has been filled in for you.

Solution:

$$
\begin{aligned}
\bar{\mathcal{L}} &= 1 && (7)\\
\bar{y} &= \bar{\mathcal{L}}[\frac{1}{y}(logy - logt)] && (8)\\
\bar{z} &= \bar{y} e^z && (9)\\
\bar{w_2} &= \bar{z} x^2 && (10)\\
& && (11)
\end{aligned}
$$

3. Your design a multilayer DN which receives three binary-valued (i.e. 0 or 1) inputs x_1, x_2, x_3, and outputs 1 if exactly two of the inputs are 1, and outputs 0 otherwise. All of the units use a hard threshold activation function:

$$z = \begin{cases} 1 & \text{if } z \geq 0 \\ 0 & \text{if } z < 0 \end{cases}$$

   (a) Draw the computation graph for this example.

   (b) Specify weights and biases which correctly implement this function. Hint: one of the hidden units should activate if 2 or more inputs are on, and the other should activate if all of the inputs are on.