CIS 700-004 Spring 2019 **Homework 0: due January 30th, 2019**

This is a review of machine learning designed as a memory refresher;f not all concepts have been discussed in class. Feel free to consult any reference materials.

1. **[10 pts]** **Understanding Entropy**

   This problem reviews Shannon's informational entropy. Please use log_2 for all computations.

   (a) Compute the entropy of the following distribution.

   $$X = \begin{cases} 1 & \text{w.p. } 0.3 \\ 0 & \text{w.p. } 0.7 \end{cases}$$

   (b) Compute the entropy of the following distribution.

   $$X = \begin{cases} 1 & \text{w.p. } 0.6 \\ 0 & \text{w.p. } 0.4 \end{cases}$$

   (c) Compute the entropy of the following distribution.

   $$X = \begin{cases} 1 & \text{w.p. } 0.998 \\ 0 & \text{w.p. } 0.002 \end{cases}$$

   (d) Note that increasing $\Pr[X = 1]$ monotonically from 0.3 to 0.6 to 0.998 does not increase entropy monotonically. Assuming two outcomes, what distribution of X maximizes entropy?

   (e) In your own words, relate entropy to the concept of surprise (aka information content).

---

2. **[10 pts]** **Understanding Cross-Entropy**

   Suppose we have a true distribution of

   $$P = \begin{cases} 10 & \text{w.p. } 0.65 \\ 11 & \text{w.p. } 0.25 \\ 12 & \text{w.p. } 0.1 \end{cases}$$

   and an estimated distribution of

   $$Q = \begin{cases} 10 & \text{w.p. } 0.1 \\ 11 & \text{w.p. } 0.4 \\ 12 & \text{w.p. } 0.5 \end{cases}$$

(a) Compute the cross-entropy $H(P, Q)$.

(b) Suppose we obtain a better estimate Q' s.t.

$$Q' = \begin{cases} 10 & \text{w.p. } 0.6 \\ 11 & \text{w.p. } 0.3 \\ 12 & \text{w.p. } 0.1 \end{cases}$$

Compute the cross-entropy $H(P, Q')$.

(c) Is cross-entropy symmetric? i.e. if we reverse the distributions for P, Q, will we get the same cross-entropy?

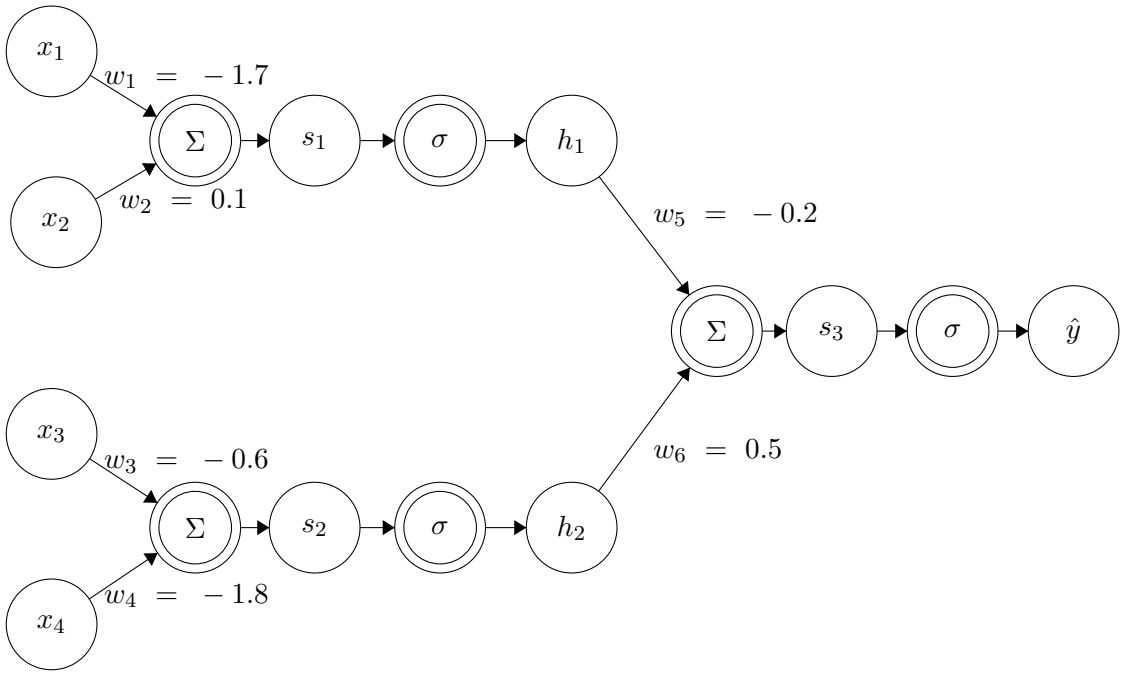
(d) Suppose we have a true label $y = [1, 0, 0]$, and a logistic regression model computes estimated probabilities $\hat{y} = [0.91, 0.08, 0.01]$. Compute the cross-entropy loss.

(e) What would happen to the cross-entropy loss if any of the estimated probabilities were zero? Is this possible in a logistic regression?

---

3. **[10 pts]** **Backpropagation by Hand**

Consider the following neural network. Single-circled nodes denote variables (e.g. x_1 is an input variable, h_1 is an intermediate variable, $\hat{y}$ is an output variable), and double-circled nodes denote functions (e.g. Σ takes the sum of its inputs, and σ denotes the logistic function $\sigma(x) = \frac{1}{1+e^{-x}}$. In the network below, $h_1 = \frac{1}{1+e^{-x_1 w_1 - x_2 w_2}}$.



Suppose we have an L2 loss $L(y, \hat{y}) = ||y - \hat{y}||_2^2$. We are given a data point $(x_1, x_2, x_3, x_4) = (-0.7, 1.2, 1.1, -2)$ with true label 0.5. Use the backpropagation algorithm to compute the partial derivative $\frac{\partial L}{w_1}$.

(Hint: the gradient of an L2 loss function $||\hat{y} - y||_2^2$ is $2||\hat{y} - y||$.)

---

4. **[10 pts]** **Other Definitions**

Define the following terms.

(a) Leave-one-out cross validation

(b) Bayes error

(c) Supervised leraning

(d) AUC

(e) bias-variance decomposition

(f) confusion matrix

(g) sparse feature

(h) PCA

(i) epoch

(j) L1 regularization