CS 4780/5780 Homework 9 Solution

Problem 1: Convolutional-network

So here we construct Kernalization $\phi(x)$, which is a clever way to make inner products computationally tractable, so

$$\phi(x) = \begin{bmatrix} h_1(x) \\ . \\ . \\ . \\ h_m(x) \end{bmatrix}$$

Each $h_i(x)$ is a linear classifier. (Not required)Here we draw the relationship among input $\vec{x}$, hidden layer $h_m(\vec{x})$ and output $H(\vec{x})$. Please see Figure 1

Problem 2: Convolutional Process

1. The dimension is $\frac{(n+2\times p-k)}{s}+1$
2. The output is

$$\begin{bmatrix} 3 & 3 \\ 7 & 0 \end{bmatrix}$$

Problem 3: RELU-network

(a) To find the decision boundary, we will first express $\sigma(z)$ in terms of x_1, x_2. Using the weights and architecture given:

$$\begin{aligned} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} &= \begin{bmatrix} 1 & -1 & 0 \\ -1 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} \\ h_1 &= x_1 - x_2 \\ h_2 &= -x_1 - x_2 \end{aligned}$$

$$\begin{aligned} z &= \begin{bmatrix} 1 & 1 & -2 \end{bmatrix} \begin{bmatrix} f(h_1) \\ f(h_2) \\ 1 \end{bmatrix} \\ &= f(h_1) + f(h_2) - 2 \end{aligned}$$

Putting this all together, we see that:

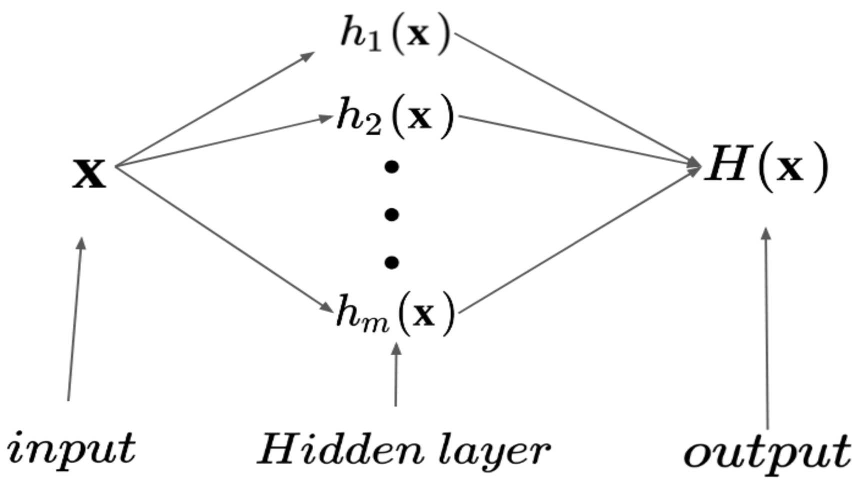

Figure 1: Neural Network

$$
\begin{aligned}
t &= \sigma(z) \\
&= \sigma(f(h_1) + f(h_2) - 2) \\
&= \sigma(f(x_1 - x_2) + f(-x_1 - x_2) - 2) \\
&= \frac{1}{1 + e^{-(max(0, x_1 - x_2) + max(0, -x_1 - x_2) - 2)}}
\end{aligned}
$$

If you do this by hand, you can split the equation into the cases in which:

- $x_1 > x_2 \ and \ x_1 + x_2 > 0$
- $x_1 > x_2 \ and \ x_1 + x_2 < 0$
- $x_1 < x_2 \ and \ x_1 + x_2 > 0$
- $x_1 < x_2 \ and \ x_1 + x_2 < 0$

For all of these, you get the following equations:

- $\frac{1}{1+e^{-(x_1 - x_2 - 2)}}$
- $\frac{1}{1+e^{2(x_2+1)}}$
- $\frac{1}{1+e^{2}}$
- $\frac{1}{1+e^{x_1+x_2+2}}$

Graphing this with Wolfram Alpha gives us:

Input:

$$0.5 = \frac{1}{1 + e^{-(\max(0, x-y) + \max(0, -x-y) - 2)}}$$

Implicit plot:

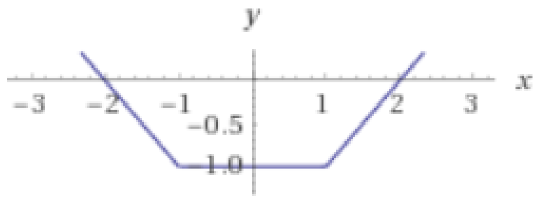

Figure 2: Decision Boundary for the RELU Network

The top is classified negative, and the bottom is positive. (See Part B for more information.)

(b) Plugging values into the equation from Part A, we see that the prediction for the point $\begin{bmatrix} 1 & 1 \end{bmatrix}^T$ is $\frac{1}{1+e^2} \approx 0.11920$. Since this is less than the decision boundary of 0.5, we classify this point as negative.

(c) Notice that $\frac{\partial t}{\partial z} = t(1-t)$

For i=1,2:

$$\frac{\partial l}{\partial v_i} = \frac{\partial l}{\partial t} \cdot \frac{\partial t}{\partial z} \cdot \frac{\partial z}{\partial v_3} = (\frac{-y}{t} + \frac{1-y}{1-t}) \cdot (t \cdot (1-t)) \cdot f(h_i) = (t-y) \cdot f(h_i)$$

For i=3:

$$\frac{\partial l}{\partial v_3} = \frac{\partial l}{\partial t} \cdot \frac{\partial t}{\partial z} \cdot \frac{\partial z}{\partial v_3} = (\frac{-y}{t} + \frac{1-y}{1-t}) \cdot (t \cdot (1-t)) \cdot 1 = t - y$$

For j=1,2:

$$\begin{aligned} \frac{\partial l}{\partial w_{ij}} &= \frac{\partial l}{\partial t} \cdot \frac{\partial t}{\partial z} \cdot \frac{\partial z}{\partial f(h_i)} \cdot \frac{\partial f(h_i)}{\partial h_i} \cdot \frac{\partial h_i}{\partial w_{ij}} \\ &= (\frac{-y}{t} + \frac{1-y}{1-t}) \cdot (t \cdot (1-t)) \cdot v_i \cdot I(h_i > 0) \cdot x_j \\ &= (t-y) \cdot v_i \cdot I(h_i > 0) \cdot x_j \end{aligned}$$

For j=3:

$$\begin{aligned} \frac{\partial l}{\partial w_{i3}} &= \frac{\partial l}{\partial t} \cdot \frac{\partial t}{\partial z} \cdot \frac{\partial z}{\partial f(h_i)} \cdot \frac{\partial f(h_i)}{\partial h_i} \cdot \frac{\partial h_i}{\partial w_{i3}} \\ &= (\frac{-y}{t} + \frac{1-y}{1-t}) \cdot (t \cdot (1-t)) \cdot I \cdot I(h_i > 0) \cdot 1 \\ &= (t-y) \cdot v_i \cdot I(h_i > 0) \end{aligned}$$