ECE 6143 Homework 2

Qianyong Tang

October 1, 2019

1 Problem 1

Since $y_i \in \{-1, 1\}$, denote $y_i^{'} = \frac{1}{2}(1 - y_i) \in \{0, 1\}$. Choose $x_j = 10^{-j}$ (where $i, j = 1, 2, 3, ..., h$), and $\omega = \pi\big(1 + \sum_{i=1}^{h} y_i^{'} 10^i\big)$

$$
\begin{aligned}
\omega x_j &= \pi\big(1 + \sum_{i=1}^{h} y_i^{'} 10^i\big) 10^{-j} \\
&= \pi\big(10^{-j} + \sum_{i=1}^{h} y_i^{'} 10^{i-j}\big) \\
&= \pi\big(10^{-j} + \sum_{i=1}^{j-1} y_i^{'} 10^{i-j} + y_i^{'} + \sum_{i=j+1}^{h} y_i^{'} 10^{i-j}\big) \\
&= \pi\big(10^{-j} + \sum_{i=1}^{j-1} y_i^{'} 10^{i-j} + y_i^{'}\big) + 10\pi \sum_{i=j+1}^{h} y_i^{'} 10^{i-(j+1)}
\end{aligned}
$$

Since $10\pi \sum_{i=j+1}^{h} y_i^{'} 10^{i-(j+1)} = 2k\pi$, we can omit this part for $f(x) = \text{sign}(\sin(\omega x))$, then

$$
\begin{aligned}
\omega x_j &= \pi\big(10^{-j} + \sum_{i=1}^{j-1} y_i^{'} 10^{i-j} + y_i^{'}\big) \\
&= \pi\big(10^{-j} + 10^{-j} \sum_{i=1}^{j-1} y_i^{'} 10^{i} + y_i^{'}\big) \\
&\leqslant \pi\big(10^{-j} + 10^{-j} \sum_{i=1}^{j-1} 10^{i} + y_i^{'}\big) \\
&< \pi\big(10^{-j} + 10^{-j}(10^j - 1) + y_i^{'}\big) \\
&= \pi\big(1 + y_i^{'}\big)
\end{aligned}
$$

Since $10^{-j} + \sum_{i=1}^{j-1} y_i^{'} 10^{i-j} > 0$, it is obvious that $\omega x_j > \pi y_i^{'}$, so $\pi y_i^{'} < \omega x_j < \pi\big(1 + y_i^{'}\big)$
For $y_i = -1 \Rightarrow y_i^{'} = 1$:

$$\begin{aligned}&\pi < \omega x_j < 2\pi \\ &f(x_j) = \text{sign}(\sin(\omega x_j)) = -1\end{aligned}$$

For $y_i = 1 \Rightarrow y_i^{'} = 0$:

$$\begin{aligned}&0 < \omega x_j < \pi \\ &f(x_j) = \text{sign}(\sin(\omega x_j)) = 1\end{aligned}$$

Since no matter what the value of h is, the classifier can always classifies correctly, so the classifiers operating in one dimension with input space X= $[0, 2\pi]$ is infinite.

2 Problem 2

2.1

First, we derive $\frac{\partial E}{\partial w_{ji}}$, which can be expressed by the chain rule:

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial x_i}\frac{\partial x_i}{\partial s_i}\frac{\partial s_i}{\partial w_{ji}}$$

We find the value of each part separately

$$\frac{\partial E}{\partial x_i} = (-1)\big(\frac{t_i}{x_i} + \frac{1 - t_i}{1 - x_i}(-1)\big) = \frac{x_i - t_i}{x_i(1 - x_i)}$$

$$\frac{\partial x_i}{\partial s_i} = (-1)(1 + e^{-s_i})^{-2} e^{-s_i}(-1) = x_i(1 - x_i)$$

$$\frac{\partial s_i}{\partial x_{ji}} = y_i$$

So,

$$\frac{\partial E}{\partial w_{ji}} = (x_i - t_i)y_i$$

Then, we derive $\frac{\partial E}{\partial w_{kj}}$. Similarly, applying chain rule and the equations obtained above, we get

$$\begin{aligned}\frac{\partial E}{\partial w_{kj}} &= \frac{\partial E}{\partial s_j}\frac{\partial s_j}{\partial w_{kj}} \\ &= \sum_i \frac{\partial E}{\partial s_i}\frac{\partial s_i}{\partial y_j}\frac{\partial y_j}{\partial s_j}\frac{\partial s_j}{\partial w_{kj}} \\ &= \sum_i (x_i - t_i) \cdot w_{ji} \cdot y_j(1 - y_j) \cdot z_k\end{aligned}$$

Therefore, we obtain the update rule

$$w_{ji}^{p+1} = w_{ji}^{p} - \eta\frac{\partial E}{\partial w_{ji}} \qquad w_{kj}^{p+1} = w_{kj}^{p} - \eta\frac{\partial E}{\partial w_{kj}}$$

2.2

$$
\begin{aligned}
\frac{\partial E}{\partial w_{ji}} &= \sum_k \frac{\partial E}{\partial x_k}\frac{\partial x_k}{\partial w_{ji}} \\
&= \sum_k \frac{\partial E}{\partial x_k}(\sum_p \frac{\partial x_k}{\partial s_p}\frac{\partial s_p}{\partial w_{ji}}) \\
&= \sum_k (-\frac{t_k}{x_k})\big(\sum_p x_p(\delta_{kp} - x_k)\delta_{pi} y_j\big) \\
&= \sum_k (-\frac{t_k}{x_k})[x_i(\delta_{ki} - x_k)y_i] \\
&= (-\frac{t_i}{x_i})[x_i(1-x_i)y_i] + \sum_{k \neq i}(-\frac{t_k}{x_k})[x_i(-x_k)y_i] \\
&= x_i y_i \sum_k t_k - t_i y_i \\
&= y_i(x_i - t_i)
\end{aligned}
$$

Similarly, we can get $\frac{\partial E}{\partial w_{kj}}$

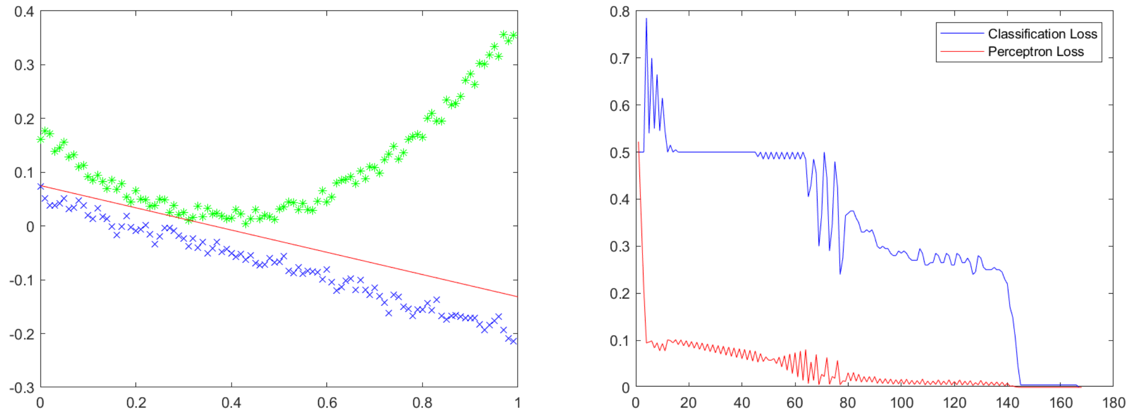
$$
\begin{aligned}
\frac{\partial E}{\partial w_{kj}} &= \sum_i \frac{\partial E}{\partial s_i} \sum_j \frac{\partial s_i}{\partial y_j} \sum_a \frac{\partial y_j}{\partial s_a}\frac{\partial s_a}{\partial w_{kj}} \\
&= \sum_i (x_i - t_i) w_{ji} y_j (1 - y_j) z_k
\end{aligned}
$$

Therefore, we obtain the update rule

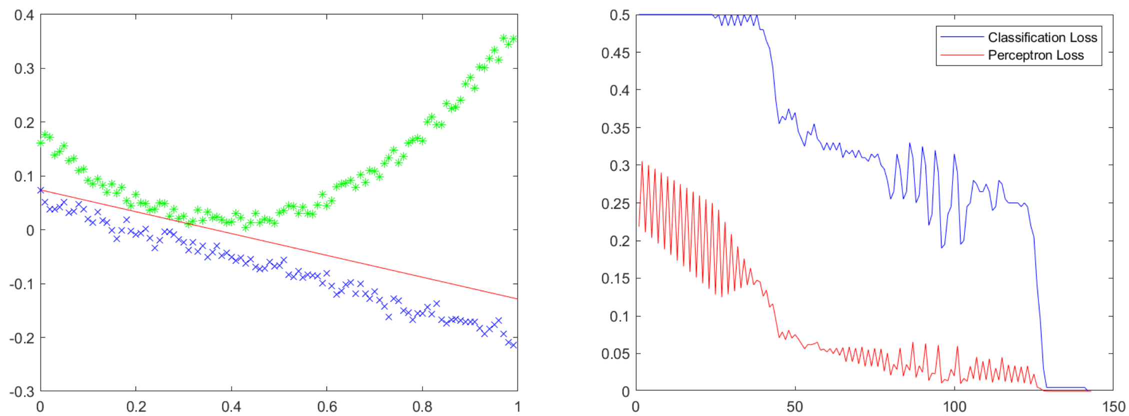
$$
w_{ji}^{p+1} = w_{ji}^p - \eta \frac{\partial E}{\partial w_{ji}} \qquad w_{kj}^{p+1} = w_{kj}^p - \eta \frac{\partial E}{\partial w_{kj}}
$$

3 Problem 3

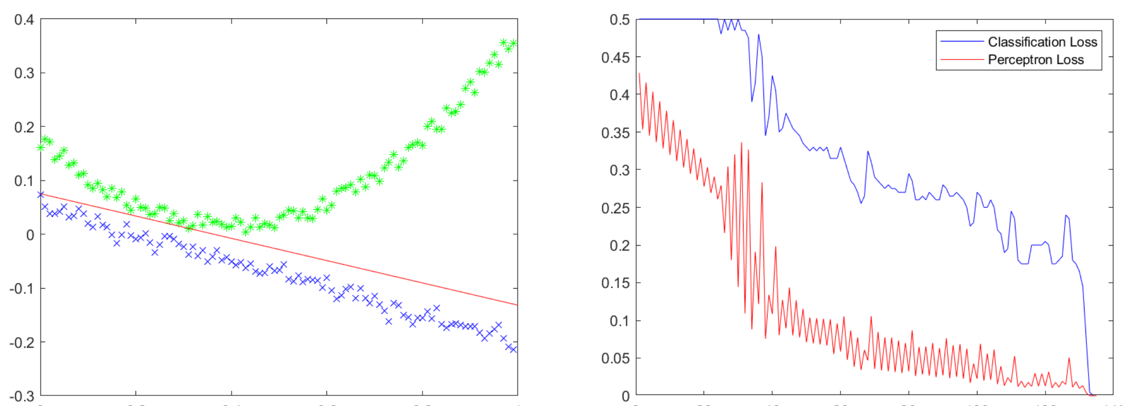
Using Gradient Descent (GD), set step = 0.5, tolerance = 0.0001



Using Gradient Descent (GD), set step = 1.5, tolerance = 0.0001



Using Gradient Descent (GD), set step = 2.5, tolerance = 0.0001



We can see that as step increases, the iteration number will decrease. However, the Classification and Perceptron Loss will become "deeper", and more fluctuation.

4 Problem 4

For the discrete distribution $\{p_k | k = 1, 2, ..., N\}$, we have $\sum_k p_k = 1$. Use the donation as follows,

$$H = -\sum_{k=1}^{N} p_k \log p_k$$
$$I = \sum_{k=1}^{N} p_k - 1$$
$$F = H + \lambda I$$

Applying Lagrange multipliers,

$$\begin{aligned}
\nabla F &= \begin{bmatrix} \frac{\partial F}{\partial p_1} \\ \frac{\partial F}{\partial p_2} \\ ... \\ \frac{\partial F}{\partial p_N} \\ \frac{\partial F}{\partial \lambda} \end{bmatrix} \\
&= \begin{bmatrix} -\log p_1 - 1 + \lambda \\ -\log p_2 - 1 + \lambda \\ ... \\ -\log p_N - 1 + \lambda \\ p_1 + p_2 + ... + p_N - 1 \end{bmatrix} \\
&= 0 \\
&\Rightarrow p_1 = p_2 = ... = p_N = e^{1 - \log N - 1} = e^{\lambda - 1} \\
&\Rightarrow \lambda = 1 - \log N \\
&\Rightarrow p_1 = p_2 = ... = p_N = e^{1 - \log N - 1} = \frac{1}{N}
\end{aligned}$$

So the uniform distribution will maximize the entropy.