Deep Learning Homework 1

Michael Remediakis

October 8, 2019

1 Perceptrons

(a) Given that $d = 2$, the perceptron activation rule can be rewritten as:

$$y = \begin{cases} 1 & \text{if } \sum_{i=0}^{2} w_i x_i \geq 0 \\ 0 & \text{else} \end{cases}$$
$$= \begin{cases} 1 & \text{if } w_2 x_2 + w_1 x_1 + w_0 \geq 0 \\ 0 & \text{else} \end{cases}$$

If we're looking for the decision boundary we simply need to find at what point $y = 0$, which occurs when our summation is less than 0, therefore we can find the decision boundary at:

$$w_2 x_2 + w_1 x_1 + w_0 = 0$$

Which we can write in slope intercept form assuming $x_2 = y(x) and x_1 = x$:

$$x_2 = -\frac{w_1}{w_2} x_1 + (-\frac{w_0}{w_2})$$
$$y(x) = -\frac{w_1}{w_2} x + (-\frac{w_0}{w_2})$$

(b) Given we can represent the discriminant function as

$$y(x) = w^T x + w_0$$

We can take a geometric representation of this equation to notice that w would be normal to the line. Therefore, if we consider x to be some point on the line, then the normal distance can be given by:

$$l = \frac{w^T x}{||w||}$$

Where $w^T = y(x) - w_0 = 0 - w_0 = w_0$ from the discriminant function, giving us:

$$l = \frac{-w_0}{||w||}$$

(c) We will define α as the learning rate, t as the teacher output, and y as the expected output. The update equation can be expressed as:

$$w_j^{(r+1)} = w_j^{(r)} + \alpha * (t - y) * x_i$$

(d)

x_1	x_2	Output	Teacher	w_0	w_1	w_2
0	1	1	0	0	0	0
0	0	0	1	-1	0	-1
1	0	1	0	0	0	-1
0	0	0	1	-1	-1	-1
0	0	1	1	0	-1	-1

(e) No, the solution is not unique, a weight vector of (0, -2, -2) would also be a sufficient perceptron that satisfies the truth table.

2 Logistic Regression

$$
\begin{aligned}
E(w) &= -\sum_{n=1}^{N} \{t^n \ln(y^n) + (1-t^n)\ln(1-y^n)\} \\
&= -\sum_{n=1}^{N} \left\{t^n \ln\left(\frac{1}{1+e^{-w_j x_j^n}}\right) + (1-t^n)\ln\left(1 - \frac{1}{1+e^{-w_j x_j^n}}\right)\right\} \\
&= -\sum_{n=1}^{N} \left\{t^n \ln 1 - t^n \ln\left(1+e^{-w_j x_j^n}\right) + (1-t^n)\ln\left(\frac{1+e^{-w_j x_j^n}-1}{1+e^{-w_j x_j^n}}\right)\right\} \\
&= -\sum_{n=1}^{N} \left\{t^n \ln 1 - t^n \ln\left(1+e^{-w_j x_j^n}\right) + (1-t^n)\ln\left(\frac{e^{-w_j x_j^n}}{1+e^{-w_j x_j^n}}\right)\right\} \\
&= -\sum_{n=1}^{N} \left\{t^n \ln 1 - t^n \ln\left(1+e^{-w_j x_j^n}\right) + (1-t^n)\ln e^{-w_j x_j^n} - (1-t^n)\ln 1+e^{-w_j x_j^n}\right\} \\
&= -\sum_{n=1}^{N} \left\{t^n \ln 1 + (1-t^n)\ln e^{-w_j x_j^n} - \ln\left(1+e^{-w_j x_j^n}\right)\right\} \\
&= -\sum_{n=1}^{N} \left\{t^n \ln 1 - (1-t^n)w_j x_j^n - \ln\left(1+e^{-w_j x_j^n}\right)\right\} \\
&= -\sum_{n=1}^{N} \left\{t^n \ln 1 - w_j x_j^n + t^n w_j x_j^n - \ln\left(1+e^{-w_j x_j^n}\right)\right\} \\
\frac{\partial E(w)}{\partial w_j} &= -\sum_{n=1}^{N} \left\{-x_j^n + t^n x_j^n - \frac{1}{1+e^{-w_j x_j^n}} * e^{-w_j x_j^n} * -x_j^n\right\} \\
&= -\sum_{n=1}^{N} \left\{(t^n-1)x_j^n + \frac{e^{-w_j x_j^n}}{1+e^{-w_j x_j^n}} * x_j^n\right\} \\
&= -\sum_{n=1}^{N} \left\{\left(t^n + \frac{e^{-w_j x_j^n}}{1+e^{-w_j x_j^n}} - 1\right) x_j^n\right\} \\
&= -\sum_{n=1}^{N} \left\{\left(t^n + \frac{e^{-w_j x_j^n} - (1+e^{-w_j x_j^n})}{1+e^{-w_j x_j^n}}\right) x_j^n\right\} \\
&= -\sum_{n=1}^{N} \left\{\left(t^n - \frac{1}{1+e^{-w_j x_j^n}}\right) x_j^n\right\} \\
&= -\sum_{n=1}^{N} (t^n - y^n)\, x_j^n \\
-\frac{\partial E(w)}{\partial w_j} &= \sum_{n=1}^{N} (t^n - y^n)\, x_j^n
\end{aligned}
$$