

ARTIFICIAL NEURAL NETWORKS

Exercise Set 1

Antonio Peters

February 24, 2016

1 Exercise 1.1.1

1.1 [1] (a)

$$f(x) = ax^2 + bx + c$$

$$\text{Let } \bar{v} = [ax_1^2 + bx_1 + c, \dots, ax_n^2 + bx_n + c]$$

$$\text{Let } \bar{y} = [y_1, y_2, \dots, y_n]$$

$$\text{Minimize } \bar{e} = \bar{v} - \bar{y} = [(ax_1^2 + bx_1 + c - y_1), \dots, (ax_n^2 + bx_n + c - y_n)]$$

We need to minimize $\|\bar{e}\|$ or equivalently minimize $S = \|\bar{e}\|^2$.

Therefore $S = \sum_{i=1}^n (x_i^2 + bx_i + c - y_i)^2$, the sum of square errors needs to be minimized.

Set:

$$\begin{aligned} \frac{\partial S}{\partial a} &= \sum_{i=1}^n 2(ax_i^2 + bx_i + c - y_i)x_i^2 = 0 \\ \frac{\partial S}{\partial b} &= \sum_{i=1}^n 2(ax_i^2 + bx_i + c - y_i)x_i = 0 \\ \frac{\partial S}{\partial c} &= \sum_{i=1}^n 2(ax_i^2 + bx_i + c - y_i) = 0 \end{aligned} \tag{1}$$

Which is equivalent to:

$$\begin{aligned}
\left(\sum_{i=1}^n x_i^4\right)a + \left(\sum_{i=1}^n x_i^3\right)b + \left(\sum_{i=1}^n x_i^2\right)c &= \sum_{i=1}^n x_i^2 y_i \\
\left(\sum_{i=1}^n x_i^3\right)a + \left(\sum_{i=1}^n x_i^2\right)b + \left(\sum_{i=1}^n x_i\right)c &= \sum_{i=1}^n x_i y_i \\
\left(\sum_{i=1}^n x_i^2\right)a + \left(\sum_{i=1}^n x_i\right)b + \sum_{i=1}^n c &= \sum_{i=1}^n y_i
\end{aligned} \tag{2}$$

Which becomes:

$$\begin{bmatrix} \sum_{i=1}^n x_i^4 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i & n \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n x_i^2 y_i \\ \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{bmatrix} \tag{3}$$

Which can be represented by:

$$A\bar{b} = C \tag{4}$$

And our coefficients can be found by solving:

$$\bar{b} = C/A \tag{5}$$

1.2 [1] (b)

See q1.1.b.m and quadreg.m

1.3 [2] (a)

$$f(x) = M(1 - e^{-kx})$$

$$\text{Let } \bar{v} = [M(1 - e^{-kx_1}), \dots, M(1 - e^{-kx_n})]$$

$$\text{Let } \bar{y} = [y_1, y_2, \dots, y_n]$$

$$\text{Minimize } \bar{e} = \bar{v} - \bar{y} = [(M(1 - e^{-kx_1}) - y_1), \dots, (M(1 - e^{-kx_n}) - y_n)]$$

We need to minimize $\|\bar{e}\|$ or equivalently minimize $S = \|\bar{e}\|^2$.

Therefore $S = \sum_{i=1}^n (M(1 - e^{-kx_i}) - y_i)^2$, the sum of square errors needs to be minimized.

Set:

$$\begin{aligned}
\frac{\partial S}{\partial M} &= \sum_{i=1}^n 2(M(1 - e^{-kx_i}) - y_i)(1 - e^{-kx_i}) = 0 \\
\frac{\partial S}{\partial k} &= \sum_{i=1}^n 2(M(1 - e^{-kx_i}) - y_i)Mx_i e^{-kx_i} = 0
\end{aligned} \tag{6}$$

Which is equivalent to:

$$\begin{aligned} \sum_{i=1}^n (M - Me^{-kx_i} - y_i)(1 - e^{-kx_i}) &= 0 \\ \sum_{i=1}^n (M - Me^{-kx_i} - y_i)Mx_ie^{-kx_i} &= 0 \end{aligned} \quad (7)$$

Which can be simplified to:

$$\begin{aligned} \sum_{i=1}^n M(1 - 2e^{-kx_i} - e^{-2kx_i}) &= y_ie^{-kx_i} + y_i \\ \sum_{i=1}^n M - Me^{-kx_i} &= y_i \end{aligned} \quad (8)$$

M and k are clearly not separable to separate the equations into a set of matrix equations, therefore, there is no explicit solution.

1.4 [2] (b and c)

See q1.2.b.m

2

See q2.m and r2.m

$$q1.1.b.m \rightarrow r2 = 0.996$$

$$q1.2.b.m \rightarrow r2 = 1$$

3 Exercise 1.6.1

See layer1.m

4 Exercise 2.1.1

4.1 [1]

We seek:

$$\begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \quad (9)$$

Therefore, we need:

$$\begin{aligned}
(0,0) &\rightarrow 0 \\
(0,1) &\rightarrow 0 \\
(1,0) &\rightarrow 0 \\
(1,1) &\rightarrow 1
\end{aligned} \tag{10}$$

Which can be derived to:

$$\begin{aligned}
hardlim(W \begin{pmatrix} 0 \\ 0 \end{pmatrix} + b) &= 0 \\
hardlim(W \begin{pmatrix} 0 \\ 1 \end{pmatrix} + b) &= 0 \\
hardlim(W \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b) &= 0 \\
hardlim(W \begin{pmatrix} 1 \\ 1 \end{pmatrix} + b) &= 1
\end{aligned} \tag{11}$$

or:

$$\begin{aligned}
W \begin{pmatrix} 0 \\ 0 \end{pmatrix} + b &< 0 \\
W \begin{pmatrix} 0 \\ 1 \end{pmatrix} + b &< 0 \\
W \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b &< 0 \\
W \begin{pmatrix} 1 \\ 1 \end{pmatrix} + b &> 0
\end{aligned} \tag{12}$$

Assuming a Weighting of (1,1), any bias between 1 and 2 will give us the needed result, in this case, a bias of 1.1 was used. See q4.m for results.

4.2 [2]

We seek:

$$\begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \rightarrow [1 \quad 0 \quad 0 \quad 1] \tag{13}$$

Therefore, we need:

$$\begin{aligned}
(0,0) &\rightarrow 1 \\
(0,1) &\rightarrow 0 \\
(1,0) &\rightarrow 0 \\
(1,1) &\rightarrow 1
\end{aligned} \tag{14}$$

Which can be derived to:

$$\begin{aligned}
hardlim(W \begin{pmatrix} 0 \\ 0 \end{pmatrix} + b) &= 1 \\
hardlim(W \begin{pmatrix} 0 \\ 1 \end{pmatrix} + b) &= 0 \\
hardlim(W \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b) &= 0 \\
hardlim(W \begin{pmatrix} 1 \\ 1 \end{pmatrix} + b) &= 1
\end{aligned} \tag{15}$$

or:

$$\begin{aligned}
W \begin{pmatrix} 0 \\ 0 \end{pmatrix} + b &> 0 \\
W \begin{pmatrix} 0 \\ 1 \end{pmatrix} + b &< 0 \\
W \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b &< 0 \\
W \begin{pmatrix} 1 \\ 1 \end{pmatrix} + b &> 0
\end{aligned} \tag{16}$$

Again assuming a weighting of (1, 1) there does not exist a bias which will produce this result, this is due to the fact that the problem is not linearly separable, meaning that a straight line does not exist which can separate the data.

5

5.1 (a)

$$\begin{aligned}
a = f(n) = \tanh(n) &= \frac{\sinh(n)}{\cosh(n)} \\
&= \frac{\frac{e^n - e^{-n}}{2}}{\frac{e^n + e^{-n}}{2}} = \frac{e^n - e^{-n}}{e^n + e^{-n}}
\end{aligned} \tag{17}$$

The derivative is therefore:

$$\dot{a} = \dot{f}(n) = \frac{d}{dn} \frac{e^n - e^{-n}}{e^n + e^{-n}} \tag{18}$$

By the quotient rule:

$$\frac{d}{dx} \left(\frac{f(x)}{g(x)} \right) = \frac{\frac{d}{dx} f(x) g(x) - f(x) \frac{d}{dx} g(x)}{g^2(x)} \tag{19}$$

This equates to:

$$\begin{aligned}
& \frac{(e^n + e^{-n})(e^n + e^{-n}) - (e^n - e^{-n})(e^n - e^{-n})}{(e^n + e^{-n})^2} \\
&= \frac{(e^n + e^{-n})^2 - (e^n - e^{-n})^2}{(e^n + e^{-n})^2} \\
&= \frac{(e^n + e^{-n})^2}{(e^n + e^{-n})^2} - \frac{(e^n - e^{-n})^2}{(e^n + e^{-n})^2} \\
&= 1 - a^2
\end{aligned} \tag{20}$$

5.2 (b)

$$a = f(n) = \text{logsig}(n) = \frac{1}{1 + e^{-n}} \tag{21}$$

The derivative is therefore:

$$\begin{aligned}
\dot{a} &= \dot{f}(n) = \frac{d}{dn} \frac{1}{1 + e^{-n}} \\
&= \left(-\frac{1}{1 + e^{-n}} \right) (-e^{-n}) \\
&= \frac{e^{-n}}{(1 + e^{-n})^2} \\
&= \frac{1 + e^{-n} - 1}{(1 + e^{-n})^2} \\
&= \frac{1 + e^{-n}}{(1 + e^{-n})^2} - \frac{1}{(1 + e^{-n})^2} \\
&= \frac{1}{1 + e^{-n}} - \frac{1}{(1 + e^{-n})^2} \\
&= a - a^2 = (1 - a)a
\end{aligned} \tag{22}$$

5.3 (c)

For a:

```
a=@(n) tansig(n);
dot_f = 1 - a(n)^2;
```

For b:

```
a=@(n) logsig(n);
dot_f = (1 - a(n))*a(n);
```