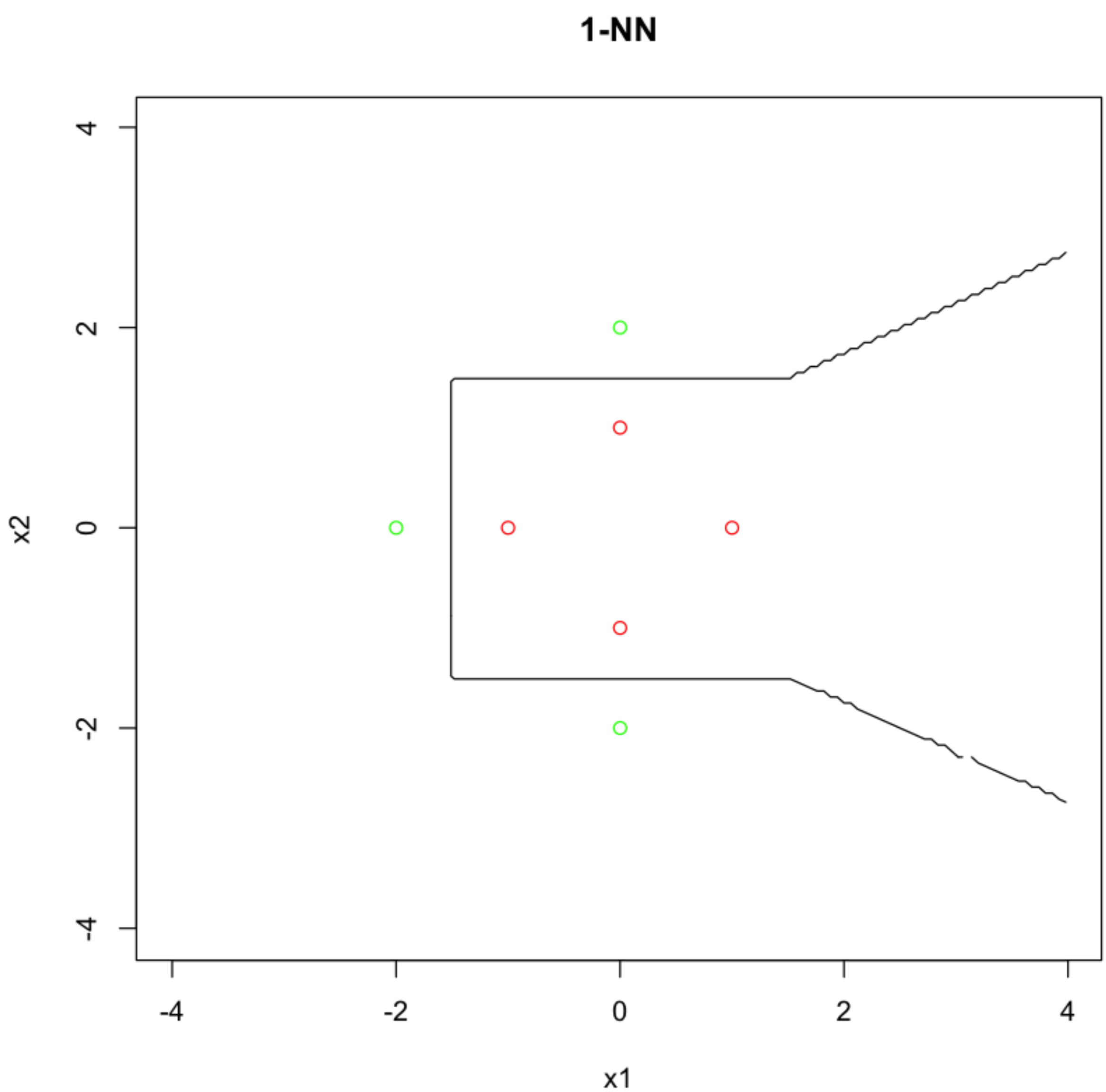
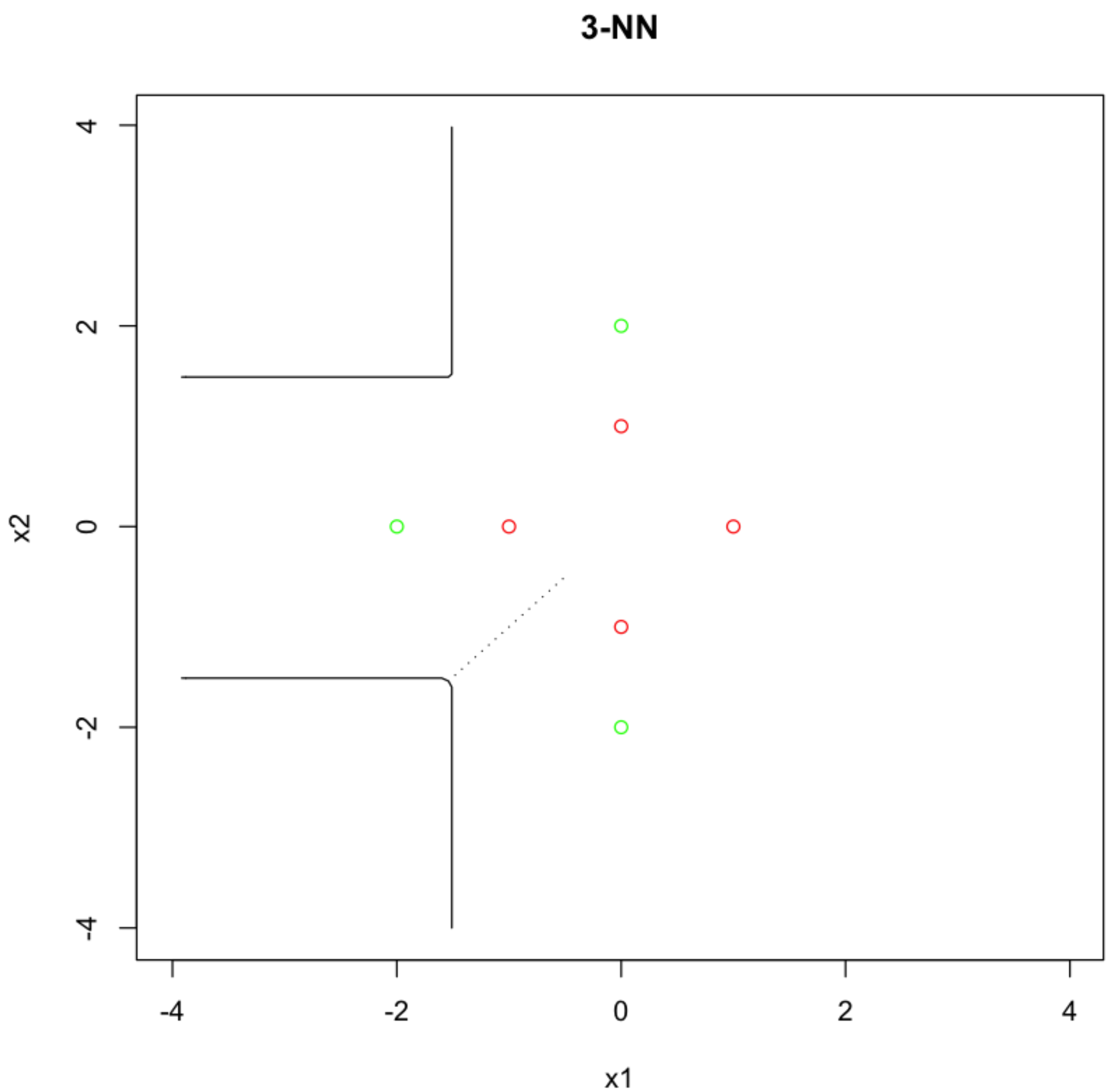
In [3]:
```
install.packages("class")
library(class)
set.seed(10)
X <- matrix(c(1, 0, 0, -1, 0, 0, -2, 0, 1, -1, 0, 2, -2, 0), nrow = 7)
seq <- seq(-4, 4, by = 0.06)
Xnew <- expand.grid(seq, seq)
labels <- c(-1, -1, -1, -1, 1, 1, 1)
NN_1 <- knn(X, Xnew, labels, k = 1, prob = TRUE)
prob <- attr(NN_1, "prob")
prob <- ifelse(NN_1 == "1", prob, 1-prob)
prob1 <- matrix(prob, length(seq), length(seq))
contour(seq, seq, prob1, levels = 0.5, labels = "", xlab = "x1", ylab = "x2
main = "1-NN")
points(X, col = ifelse(labels == -1, "red", "green"))
```

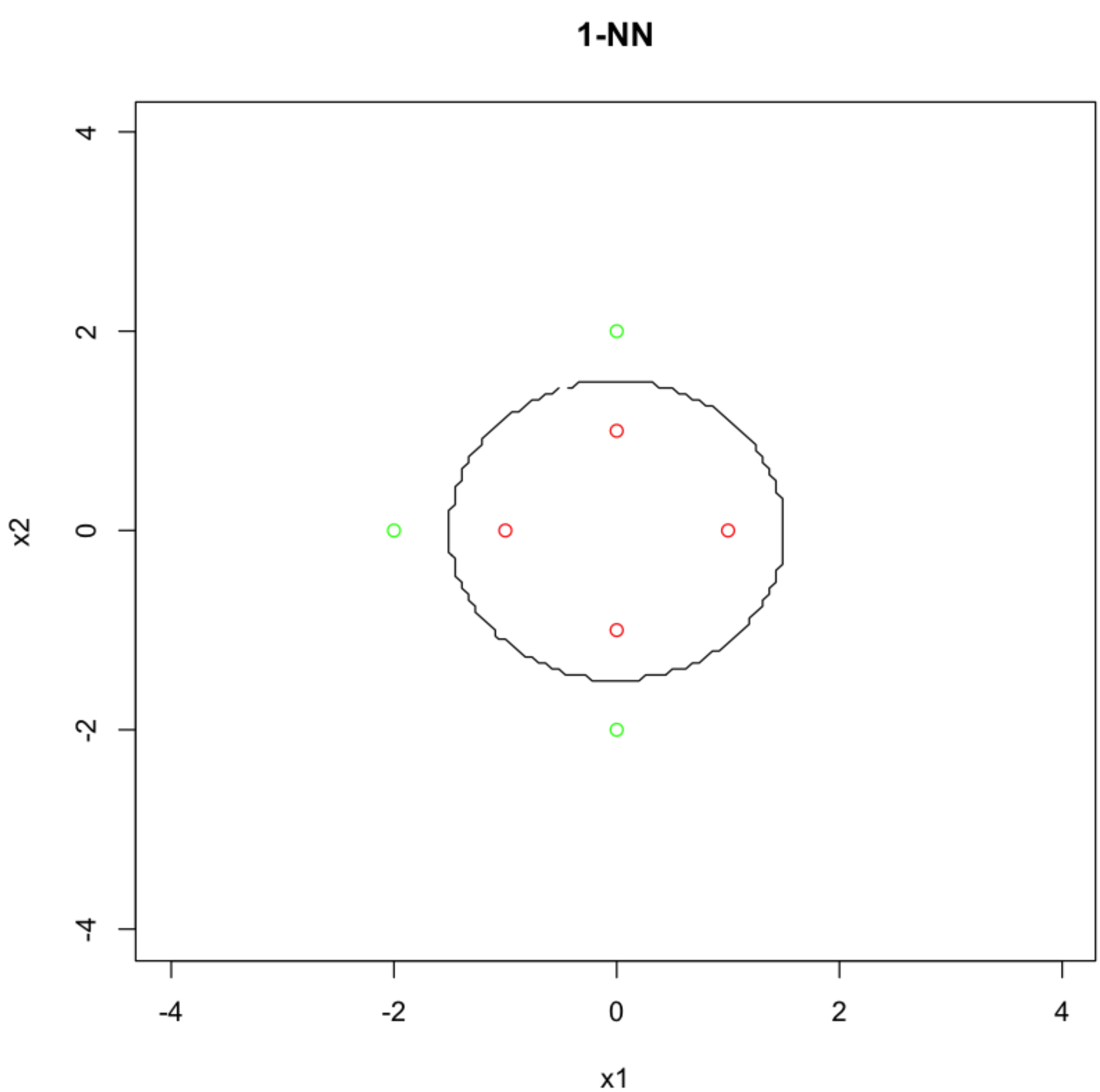
```
The downloaded binary packages are in
        /var/folders/0f/6n5nq50962qf69ffsbjxzy7r0000gn/T//RtmpETXHQ8/down
loaded_packages
```



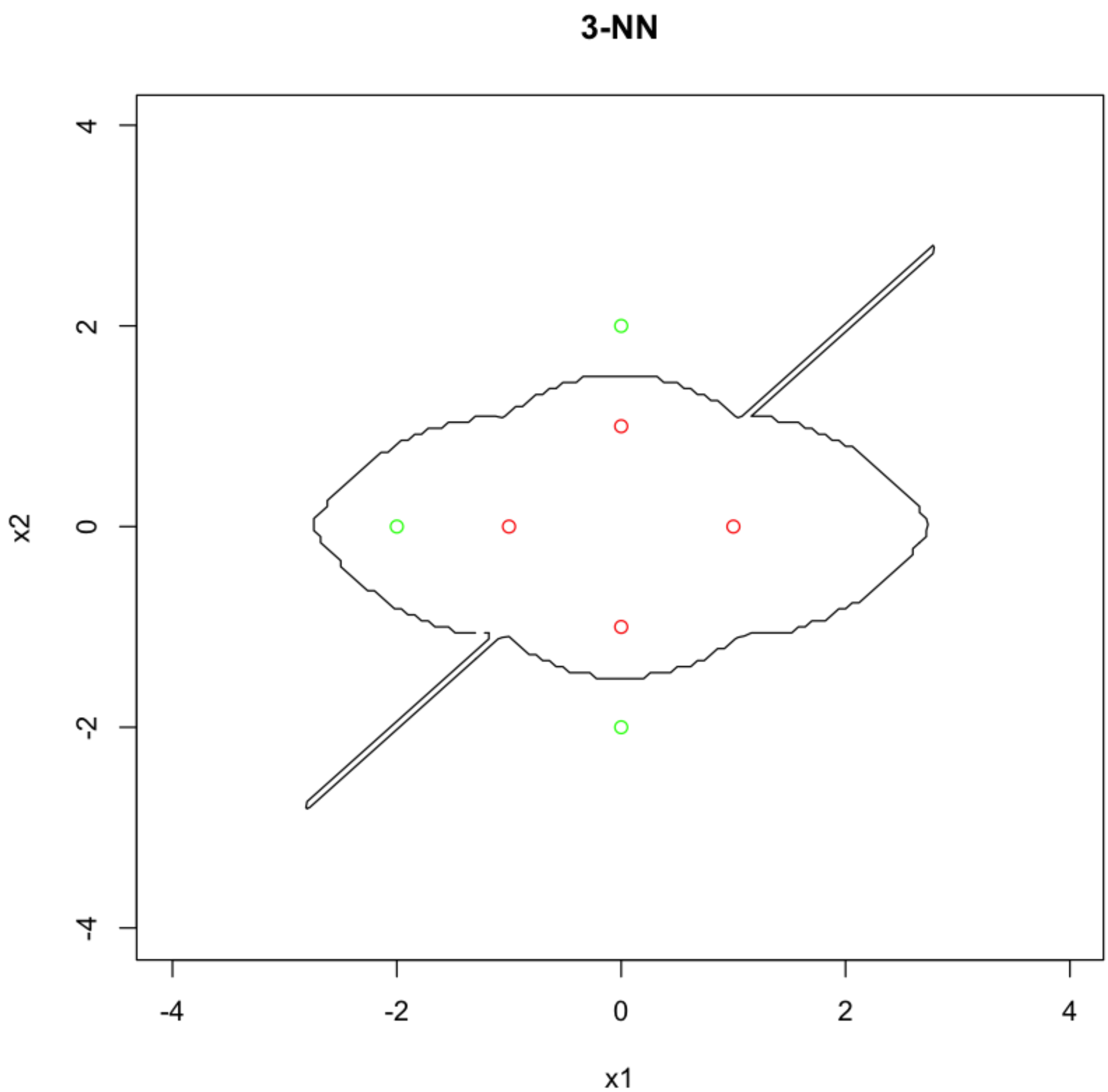
In [4]:
```
NN_3 <- knn(X, Xnew, labels, k = 3, prob = TRUE)
prob <- attr(NN_3, "prob")
prob <- ifelse(NN_3 == "1", prob, 1-prob)
prob1 <- matrix(prob, length(seq), length(seq))
contour(seq, seq, prob1, levels = 0.5, labels = "", xlab = "x1", ylab = "x2
main = "3-NN")
points(X, col = ifelse(labels == -1, "red", "green"))
```



In [5]:
```
Z1 <- apply(X, 1, function(x) sqrt(x[1]^2 + x[2]^2))
Z2 <- apply(X, 1, function(x) atan(x[2]/ x[1]))
Z <- matrix(c(Z1, Z2), byrow = FALSE, ncol = 2)
Znew1 <- apply(Xnew, 1, function(x) sqrt(x[1]^2 + x[2]^2))
Znew2 <- apply(Xnew, 1, function(x) atan(x[2]/ x[1]))
Znew <- matrix(c(Znew1, Znew2), byrow = FALSE, ncol = 2)
NN_1 <- knn(Z, Znew, labels, k = 1, prob = TRUE)
prob <- attr(NN_1, "prob")
prob <- ifelse(NN_1 == "1", prob, 1-prob)
prob1 <- matrix(prob, length(seq), length(seq))
contour(seq, seq, prob1, levels = 0.5, labels = "", xlab = "x1", ylab = "x2
main = "1-NN")
points(X, col = ifelse(labels == -1, "red", "green"))
```



In [6]:
```
NN_3 <- knn(Z, Znew, labels, k = 3, prob = TRUE)
prob <- attr(NN_3, "prob")
prob <- ifelse(NN_3 == "1", prob, 1-prob)
prob1 <- matrix(prob, length(seq), length(seq))
contour(seq, seq, prob1, levels = 0.5, labels = "", xlab = "x1", ylab = "x2
main = "3-NN")
points(X, col = ifelse(labels == -1, "red", "green"))
```



In [ ]:

Exercise 7.3

Equivalence relation $\Rightarrow h_m(x) = c_m$

$\Leftrightarrow h_m^{c_m}(x) = +1.$

if $c_m = +1.$

$$h_m^{c_m}(x) = h_m(x) = c_m = +1.$$

if $c_m = -1.$

$$h_m^{c_m}(x) = \bar{h}_m(x) = \bar{c}_m = +1.$$

condition is sufficient if $c_m = +1.$

$$+1 = h_m^{c_m}(x) = h_m(x).$$

$\therefore h_m(x) = +1 = c_m.$

and if $c_m = -1.$

$$+1 = h_m^{c_m}(x) = \bar{h}_m(x).$$

$\therefore h_m(x) = -1 = c_m.$

$\Rightarrow x \in R.$

$(h_1(x) \ldots, h_m(x)) = (c_1 c_2 \ldots c_M)$

$\Rightarrow h_m^{c_m}(x) = +1$

$\Rightarrow \prod_{m=1}^{m} h_m^{c_m}(x) = +1.$

$t_h(x) = +1.$

This means $x \notin h \Leftrightarrow t_h(x) = -1.$

if x is in a positive region $(f(x) = +1)$
then $x \in h_i$ & $t_h(x) = +1$ for all i.

Hence, $t_{h_i}(x) + \dots + t_{h_k}(x) = +1 = f(x).$

& if x is in negative region $(f(x) = -1)$
then $x \notin h_i$ & $t_{h_i}(x) = -1$ for all i.

$\therefore t_{h_i}(x) + \dots + t_{h_k}(x) = -1 = f(x).$

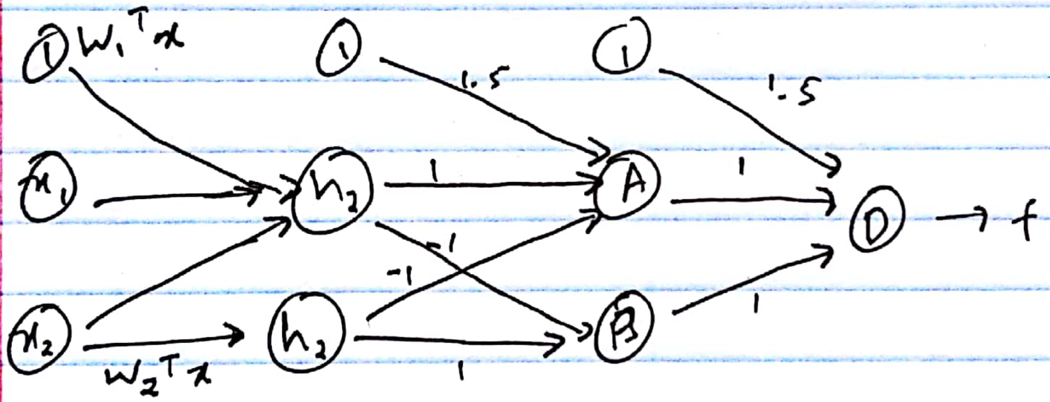
$\Rightarrow f = t_{h_1} + t_{h_2} + \dots + t_{h_k}$ where $h_1 \dots h_k$ are all
on positive region.

Excercise 7.3.

→ To show,

$f(x) = \text{sign}[\text{sign}(h_1(x) - h_2(x) - 3/2] - \text{sign}$
$[h_1(x) - h_2(x) + 3/2 + 3/2].$

Where :
$h_1(x) = \text{sign}(w_1^T x)$ and $h_2(x) = \text{sign}(w_2^T x)$



Now $A = 1.5(1) + (1)(h_1(x)) + (-1)(h_2(x))$

$\therefore (A(x)) = \text{sign}(-1.5 + h_1(x) - h_2(x))$
similarly, we can write.

$B = (-1.5)(1) + (-1)(h_1(x)) + (1)(h_2(x))$
$\therefore B(x) = \text{sign}(-1.5 - h_1(x) + (1) h_2(x))$

The output of the next layers is
output layer will become,

$(1.5)(1) \quad + (1)(A(x)) + (1)(B(x))$

$\therefore f = \text{sign}(1.5 + A(x) + B(x))$.

$\therefore$ substituting $A(x)$ & $B(x)$ in the above eqns,

$f = \text{sign}(+1.5 + \text{sign}(-1.5 + h_1(x) - h_2(x)) + \text{sign}$
$(h_2(x) - h_1(x) - 1.5)$

$\therefore f = \text{sign}[\text{sign}(h_1(x) - h_2(x) - 3/2] - \text{sign}$
$(h_1(x) - h_2(x) + 3/2) + 3/2$.

Exercise 7.7.

For the sigmoidal perceptron, $h(x) = \tan(w^T x)$, let the in-sample error be

$$E_{in}(w) = \frac{1}{N} \sum_{n=1}^{N} (\tanh(w^T x_n) - y_n)^2.$$

Show that:

$$\nabla E_{in}(w) = \frac{2}{N} \sum_{n=1}^{N} (\tanh(w^T x_n) - y_n)(1 - \tanh^2(w^T x_n)) x_n.$$

If $w \to \infty$, what happens to the gradient: how this is related to why it is hard to optimize the perceptron.

$$\nabla E_{in}(w) = \nabla_w \left(\frac{1}{N} \sum_{n=1}^{N} (\tanh(w^T x_n) - y_n)^2 \right).$$

$$= \frac{1}{N} \sum_{n=1}^{N} \nabla_w (\tanh(w^T x_n) - y_n)^2).$$

$$= \frac{1}{N} \sum_{n=1}^{N} \frac{\partial(\tanh(w^T x_n) - y_n)^2}{\partial(\tanh(w^T x_n) - y_n)} \nabla_w (\tanh(w^T x_n) - y_n).$$

$$= \frac{1}{N} \sum_{n=1}^{N} 2(\tanh(w^T x) - y_n) \nabla_w (\tanh(w^T x) - y_n).$$

$$= \frac{2}{N} \sum_{n=1}^{N} (\tanh(w^T x_n) - y_n) \nabla_w (\tanh(w^T x_n) - y_n).$$

$$= \frac{2}{N} \sum_{n=1}^{N} (\tanh(w^T x_n) - y_n) - \frac{\partial(\tanh(w^T x_n))}{\partial(w^T x)} \nabla_w (w^T x_n).$$

$$= \frac{2}{N} \sum_{n=1}^{N} (\tanh(w^T x_n) - y_n)(1 - \tanh^2(w^T x_n)) \nabla_w (w^T x_n)$$

$$= \frac{2}{N} \sum_{n=1}^{N} (\tanh(w^T x_n) - y_n)(1 - \tanh^2(w^T x_n)) x_n.$$

We see that:

$$w \to \infty \Rightarrow w^T x \to \infty \Rightarrow \begin{cases} \tan(w^T x) \to 1 \text{ if } w^T x \to +\infty \\ \tan(w^T x) \to -1 \text{ if } w^T x \to -\infty \end{cases}$$

$$\Rightarrow \tanh^2(w^T x) \to 1$$

$$\Rightarrow \tanh^2(w^T x) - 1 \to 0$$

$$\Rightarrow \nabla E_{in}(w) \to 0.$$

Scanned by CamScanner

This means that when w is large enough, the Gradient Descent algorithm will not make much change to w. Even worse, it may stop when Ein is currently at its largest possible value ($\forall (x_n, y_n) . \tanh(w^T x_n) y_n < 0$). and return that large w as the final hypothesis that it could have found.

Exercise. 7.8

a) We know $\{h_1, h_2, h_3\}$ is a V-arrangement where $h_1 < h_2 < h_3$.

$\therefore E(h_2) < \min\{E(h_1), E(h_3)\}$

$\because E(h)$ is a quadratic curve, we know its decreasing to the left of its minimum $\bar{h}$.

assuming $\bar{h} < h_1$

$$E(h_1) \leq E(h_2) \leq E(h_3).$$

is not possible as per V-arrangement.

assuming $\bar{h} > h_3$.

$$E(h_1) \geq E(h_2) \geq E(h_3).$$

is not possible as per V-arrangement.

$$\therefore \bar{h} \in [h_1, h_3].$$

(b) $e_1 = E(h_1)$, $e_2 = E(h_2)$, $e_3 = E(h_3)$.

where $e_1 = ah_1^2 + bh_1 + c$ $\quad e_2 = ah_2^2 + bh_2 + c$.

$e_3 = ah_3^2 + bh_3 + c$.

$$\therefore \quad E(h_1) = ah_1^2 + bh_1 + c = e_1$$
$$E(h_2) = ah_2^2 + bh_2 + c = e_2.$$
$$E(h_3) = ah_3^2 + bh_3 + c = e_3.$$

let D be the determinant

$$D = \begin{vmatrix} h_1^2 & h_1 & 1 \\ h_2^2 & h_2 & 1 \\ h_3^2 & h_3 & 1 \end{vmatrix}$$

$D \neq 0$ since $h_1 < h_2 < h_3$

$$\therefore \quad a = \begin{vmatrix} e_1 & h_1 & 1 \\ e_2 & h_2 & 1 \\ e_3 & h_3 & 1 \end{vmatrix} \quad \Bigg/ D = \frac{(e_1-e_2)(h_1-h_3)-(e_1-e_3)(h_1-h_2)}{D}.$$

$$b = \begin{vmatrix} h_1^2 & e_1 & 1 \\ h_2^2 & e_2 & 1 \\ h_3^2 & e_3 & 1 \end{vmatrix} \Big/ D = \frac{(e_1-e_2)(h_1^2-h_3^2) + (e_1-e_3)*(h_1^2-h_2^2)}{D}.$$

$\therefore$ minimum of such quadratic function is given by $-b/2a$.

Hence

$$\bar{h} = \frac{1}{2}\left[\frac{(e_1-e_2)(h_1^2-h_3^2)-(e_1-e_3)(h_1^2-h_2^2)}{(e_1-e_2)(h_1-h_3)-(e_1-e_3)(h_1-h_2)}\right]$$

(C) *) if $\bar{h} < h_2$

$\Rightarrow$ if $E(\bar{h}) < E(h_2)$, then $\{h_1, h_2, h_3\}$ is a new V-arrangement.

$\Rightarrow$ if $E(\bar{h}) > E(h_2)$, then $\{h_1, h_2, h_3\}$ is a V-arrangement.

*). if $\bar{h} > h_2$

$\Rightarrow$ if $E(\bar{h}) < E(h_2)$ then $\{h_1, h_2, h_3\}$ is a new V-arrangement.

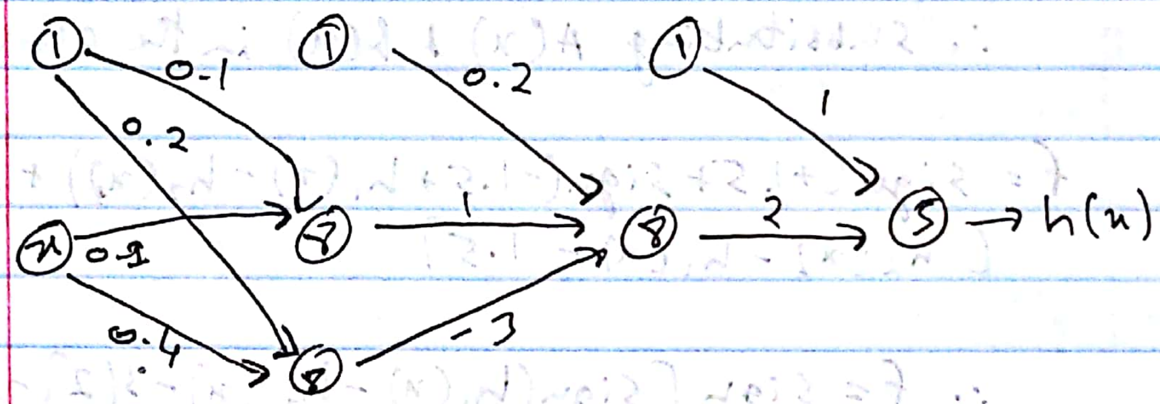
$\Rightarrow$ if $E(\bar{h}) > E(h_2)$, then $\{h_1, h_2, h_3\}$ is a new N-arrangement.

d) if $\bar{h} = h_2$, by continuity we have another h_2' close to h_2 such that $E(h_2') < \min \{E(h_1), E(h_3)\}$.

Hence we can use this h_2' in place of h_2 and proceed with the algorithm.

Exercise 7.8.

→ Given,



We get the following from above.

$$W^{(1)} = \begin{bmatrix} 0.1 & 0.2 \\ 0.3 & 0.4 \end{bmatrix} \quad W^{(2)} = \begin{bmatrix} 0.2 \\ +1 \\ -3 \end{bmatrix} \quad W^{(2)} = \begin{bmatrix} +1 \\ 2 \end{bmatrix}$$

$x = 2.$

$X^{(0)}$	$S^{(1)}$	$X^{(1)}$	$S^{(2)}$	$X^{(2)}$	$S^{(3)}$	$X^{(3)}$
$\begin{bmatrix} 1 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 0.7 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0.7 \\ 1 \end{bmatrix}$	$[-2.1]$	$\begin{bmatrix} 1 \\ -2.1 \end{bmatrix}$	$[-3.2]$	$[-3.2]$

$$\delta(3) = 2(x^3 - 1)(1)$$
$$= 2(-3.2 - 1)$$
$$= -8.4$$

$\theta'(s) = 1$

$$\delta^{(2)} = \theta'(s^{(2)}) * \left[W^{(3)} \delta^{(3)}\right]^{d(2)}$$

$$= 1 * [2 \times -8.4] = [-16.8] \rightarrow (2)$$

$$\delta^{(1)} = \theta'(s^{(1)}) * \left[W^{(2)} \delta^{(2)}\right]^{d(1)}_{1}$$

$$= 1 \times \left[\begin{pmatrix} 1 \\ -3 \end{pmatrix} [-16.8]\right]$$

$$= \begin{bmatrix} -16.8 \\ 50.4 \end{bmatrix} \rightarrow (3)$$

$$\frac{d(e)}{d(W^{(1)})} = x\,(\delta^{(1)})^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} -16.8 & 50.4 \end{bmatrix}$$

$$= \begin{bmatrix} -16.8 & 50.4 \\ -33.6 & 100.8 \end{bmatrix}$$

$$\frac{d(e)}{d(W^{(2)})} = x'\,(\delta^{(2)})^T = \begin{bmatrix} 1 \\ 0.7 \\ 1 \end{bmatrix} [-16.8]$$

$$= \begin{bmatrix} -16.8 \\ -11.76 \\ -16.8 \end{bmatrix}$$

Scanned by CamScanner

$$\frac{d(e)}{d(w^{(3)})} = x^{(2)} (\delta^{(3)})^T = \begin{bmatrix} 1 \\ -2.1 \end{bmatrix} \begin{bmatrix} -8.4 \end{bmatrix}$$

$$= \begin{bmatrix} -8.4 \\ 17.64 \end{bmatrix}$$