Q2c). Using the contour plots and surface plots, we can see that as the amount of overlap in the two clusters decreases, the decision boundary (black diagonal line) is better able to classify points being in one class or the other. In the first plot, since the two clusters overlap equally, any point on the line is equally likely to be in one of the two classes. We can also see that the contour lines are closest together near the decision boundary and get further away as there is less overlap in the two clusters. The contour lines show where the sigmoid function changes rapidly, so when there is less overlap in the clusters, the sigmoid function gets flatter and points are clearly in one class or the other. Also, as the overlap between the two clusters reduces, the accuracy of the classifier increases as the decision boundary gets more accurate and is better able to classify points being in one class or the other.

CSC411 Assignment 2

Akhil Gupta
1000357071

#4. $l(w,w_0) = -\sum_n t^{(n)} \log P(C=1|x^{(n)}) - \sum_n (1-t^{(n)}) \log P(C=0|x^{(n)})$

$P(C=0|x^{(n)}) = \frac{1}{1+e^{-z}}$; $z = w^T x + w_0$

$l(w,w_0) = \sum_n t^{(n)} z^{(n)} + \sum_n \log\left[1+e^{-z^{(n)}}\right]$

$P(C=1|x^{(n)}) = \frac{1}{1} - \frac{1}{1+e^{-z}}$

$y^{(n)} = \frac{e^{-z}}{1+e^{-z}}$

a) let $y^{(n)} = P(C=1|x^{(n)})$

show, $\frac{\partial l(w,w_0)}{\partial w_j} = \sum_n [t^{(n)} - y^{(n)}] x_j^{(n)}$

Proof: $l(w,w_0) = -\sum_n t^{(n)} \log y^{(n)} - \sum_n (1-t^{(n)}) \log(1-y^{(n)})$

$\log(1-y^{(n)}) = \log\left(\frac{1}{1+e^{-z^{(n)}}}\right) = -\log\left(1+e^{-z^{(n)}}\right)$

$\log y^{(n)} = \log\left(\frac{e^{-z^{(n)}}}{1+e^{-z^{(n)}}}\right) = -e^{-z^{(n)}} - \log\left(1+e^{-z^{(n)}}\right)$

$\therefore \frac{\partial \log y^{(n)}}{\partial w_j} = \frac{\partial\left(-e^{-z^{(n)}} - \log(1+e^{-z^{(n)}})\right)}{\partial w_j} = -x_j^{(n)} + x_j^{(n)}(1-y^{(n)}) = -y^{(n)} x_j^{(n)}$

$\frac{\partial \log(1-y^{(n)})}{\partial w_j} = \frac{\partial\left(-\log(1+e^{-z^{(n)}})\right)}{\partial w_j} = \frac{x_j^{(n)} \cdot e^{-z^{(n)}}}{1+e^{-z^{(n)}}} = x_j^{(n)}(1-y^{(n)})$

$\therefore \frac{\partial l(w,w_0)}{\partial w_j} = -\sum_n t^{(n)} \frac{\partial \log y^{(n)}}{\partial w_j} - \sum_n (1-t^{(n)}) \frac{\partial \log(1-y^{(n)})}{\partial w_j}$

$= -\sum_n t^{(n)}\left[-y^{(n)} x_j^{(n)}\right] - \sum_n (1-t^{(n)})\left[x_j^{(n)}(1-y^{(n)})\right]$

$= +\sum_n t^{(n)} y^{(n)} x_j^{(n)} - \sum_n (1-t^{(n)})\left[x_j^{(n)} - x_j^{(n)} y^{(n)}\right]$

$$= \sum_n t^{(n)} y^{(n)} x_j^{(n)} - (1-t^{(n)})[x_j^{(n)} - x_j^{(n)} y^{(n)}]$$

$$= \sum_n \left[t^{(n)} \cancel{y^{(n)} x_j^{(n)}} - x_j^{(n)} + x_j^{(n)} y^{(n)} + t^{(n)} x_j^{(n)} - t^{(n)} \cancel{x_j^{(n)} y^{(n)}} \right]$$

$$= \sum_n [x_j^{(n)} y^{(n)} - x_j^{(n)} + t^{(n)} x_j^{(n)}]$$

$$= \sum_n [y^{(n)} + t^{(n)}] - (x_j^{(n)})$$

$$= \sum_n [t^{(n)} - y^{(n)}] x_j^{(n)}$$

b) Since w_0 is the bias term, it's value is always 1 & so is $x_0^{(n)}$.

Hence, $\frac{\partial l(w, w_0)}{\partial w_0} = \sum_n [t^{(n)} - y^{(n)}] \cdot x_0^{(n)}$

$$= \sum_n [t^{(n)} - y^{(n)}] \cdot 1 = \sum_n [t^{(n)} - y^{(n)}]$$

c) To prove: $\frac{\partial l(w, w_0)}{\partial w} = X^T(t-y)$ where X = data matrix

$t = [t^{(1)}, t^{(2)}, \ldots, t^{(N)}]$

$y = [y^{(1)}, y^{(2)}, \ldots, y^{(N)}]$

$\frac{\partial l}{\partial w}$ is also column vector

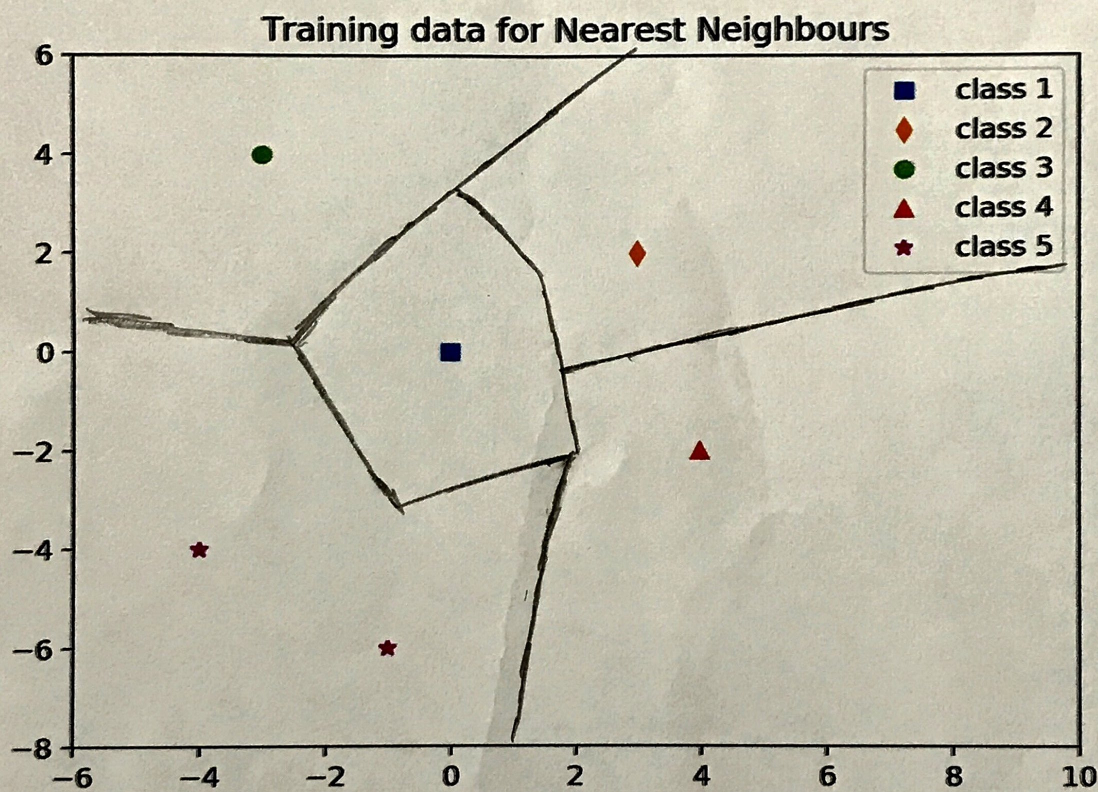
$$X = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nn} \end{pmatrix}$$

we know that $\frac{\partial l(w, w_0)}{\partial w_j}$ is a dot product of $[t^{(n)} - y^{(n)}]$ & $x_j^{(n)}$,

the equivalent row-vector translation would be $(t-y)^T X$, but since we are calculating the gradient, we would like it in column-vector form, which can be achieved by taking the transpose of $(t-y)^T X$,

and that gives us, $\frac{\partial l(w, w_0)}{\partial w} = (t-y) X^T$.

Figure 5:


Training data for Nearest Neighbours
class 1
class 2
class 3
class 4
class 5
6
4
2
0
−2
−4
−6
−8
−6
−4
−2
0
2
4
6
8
10

CSC411 Assignment 2

Akhil Gupla
1000357071

#6. Decision Trees

Entropy $= \sum_{i=1}^{c} -p_i \log_2 p_i$ $\qquad E(T,X) = \sum_{c \in X} P(c) \cdot E(c)$

$E(\text{LikeCar}) = E(3,3)$
$= E(3/6, 3/6)$
$= -\left(\frac{3}{6}\log_2\frac{3}{6}\right) - \left(\frac{3}{6}\log_2\frac{3}{6}\right) = 0.5 + 0.5 = \underline{1}$

I. Split on Make

	LikeCar	Yes	No
Make	Toyota	1	2
	Ford	2	1

$E(\text{LikeCar}, \text{Make}) = P(\text{Toyota}) \cdot E(1,2) + P(\text{Ford}) \cdot E(2,1)$

$= \frac{3}{6} \cdot E(1,2) + \frac{3}{6} \cdot E(2,1)$

$= \frac{3}{6}\left(E(1,2) + E(2,1)\right)$

$= \frac{3}{6}(1.92) = \underline{0.96}$

$E(1,2) = E(1/6, 2/6)$
$= -\left(\frac{1}{6}\log_2\frac{1}{6}\right) - \left(\frac{2}{6}\log_2\frac{2}{6}\right)$
$= \underline{0.96} = \underline{E(2,1)}$

$IG(\text{LikeCar}, \text{Make}) = E(\text{LikeCar}) - E(\text{LikeCar}, \text{Make})$
$= 1 - 0.96 = \underline{0.04}$

II. Split on Type

	LikeCar	Yes	No
Type	Sedan	3	2
	SUV	0	1

$E(\text{LikeCar}, \text{Type}) = P(\text{Sedan}) \cdot E(3,2) + P(\text{SUV}) \cdot E(0,1)$

$= \frac{5}{6} \cdot E(3,2) + \frac{1}{6} \cdot E(0,1)$

$= \frac{5}{6}(1.02) + \frac{1}{6}(0.43)$

$= \underline{0.93}$

$E(3,2) = E(3/6, 2/6)$
$= -\left(\frac{3}{6}\log_2\frac{3}{6}\right) - \left(\frac{2}{6}\log_2\frac{2}{6}\right)$
$= \underline{1.02}$

$E(0,1) = E(0/6, 1/6)$

$= -(0 \log_2 0) - (\frac{1}{6} \log_2 \frac{1}{6})$

$= \underline{0.43}$

$IG(\text{LikeCar}, \text{Type}) = E(\text{LikeCar}) - E(\text{LikeCar}, \text{Type})$

$= 1 - 0.93 = \underline{0.07}$

III. Split on colour

	LikeCar	Yes	No
Colour	Blue	2	1
	White	1	2

$E(\text{LikeCar}, \text{Color}) = P(\text{Blue}) \cdot E(2,1) + P(\text{White}) \cdot E(1,2)$

$= \frac{3}{6} \cdot E(2,1) + \frac{3}{6} \cdot E(1,2)$

$= \underline{0.96}$

$IG(\text{LikeCar}, \text{Color}) = E(\text{LikeCar}) - E(\text{LikeCar}, \text{Color})$

$= 1 - 0.96 = \underline{0.04}$

IV. Split on Cost

	LikeCar	Yes	No
Cost	High	1	1
	Low	2	2

$E(\text{LikeCar}, \text{Cost}) = P(\text{High}) \cdot E(1,1) + P(\text{Low}) \cdot E(2,2)$

$= \frac{2}{6} \cdot E(1,1) + \frac{4}{6} \cdot E(2,2)$

$= \frac{2}{6}(0.86) + \frac{4}{6}(1.05)$

$= 0.286 + 0.7 = \underline{0.986}$

$E(1,1) = E(1/6, 1/6)$

$= -(\frac{1}{6} \log_2 \frac{1}{6}) - (\frac{1}{6} \log_2 \frac{1}{6})$

$= \underline{0.86}$

$IG(\text{LikeCar}, \text{color}) = E(\text{LikeCar}) - E(\text{LikeCar}, \text{Color})$

$= 1 - 0.986$

$= \underline{0.014}$

$E(2,2) = E(2/6, 2/6)$

$= -(\frac{2}{6} \log_2 \frac{2}{6}) - (\frac{2}{6} \log_2 \frac{2}{6})$

$= \underline{1.05}$

from I, II, III, IV,

IG (LikeCar, Make) = 0.04
IG (LikeCar, Type) = 0.07
IG (LikeCar, color) = 0.04
IG (LikeCar, Cost) = 0.014

∴ hence we pick the first attribute with the highest information gain, so we split on type.

Decision Tree so far...

Type

Sedan | SUV

Make	Color	Cost	✓
Toyota	B	H	✓
Toyota	W	L	X
Ford	B	L	✓
Toyota	W	L	X
Ford	W	L	✓

Make	Color	Cost	✓
Ford	B	H	X

leaf node

$$\text{Entropy (LikeCar)} = E(3,2) = E(3/5, 2/5)$$
$$= -\left(\frac{3}{5}\log_2\frac{3}{5}\right) - \left(\frac{2}{5}\log_2\frac{2}{5}\right)$$
$$= \underline{0.97}$$

$$E(2,0) = E(2/5, 0/5)$$
$$= -\left(\frac{2}{5}\log_2\frac{2}{5}\right)$$
$$= \underline{0.53}$$

V. Split on Make

	LikeCar	Yes	No
Make	Toyota	1	2
	Ford	2	0

$$E(1,2) = E(1/5, 2/5)$$
$$= -\left(\frac{1}{5}\log_2\frac{1}{5}\right) - \left(\frac{2}{5}\log_2\frac{2}{5}\right)$$
$$= \underline{0.99}$$

$$E(\text{LikeCar, Make}) = P(\text{Toyota})\cdot E(1,2) + P(\text{Ford})\cdot E(2,0)$$
$$= \frac{3}{5}\cdot E(1,2) + \frac{2}{5}\cdot E(2,0)$$
$$= \frac{3}{5}(0.99) + \frac{2}{5}(0.53)$$
$$= 0.594 + 0.212 = \underline{0.806}$$

$$IG(\text{LikeCar, Make}) = E(\text{LikeCar}) - E(\text{LikeCar, Make})$$
$$= 0.97 - 0.806 = \underline{0.164}$$

VI. Split on Colour

	LikeCar	Yes	No
Colour	Blue	2	0
	White	1	2

$E(\text{LikeCar}, \text{Color}) = P(\text{Blue}) \cdot E(2,0) + P(\text{White}) \cdot E(1,2)$

$= \frac{2}{5} \cdot E(2,0) + \frac{3}{5} \cdot E(1,2)$

$= \frac{2}{5}(0.53) + \frac{3}{5}(0.99)$

$= 0.806$

$IG(\text{LikeCar}, \text{Color}) = E(\text{LikeCar}) - E(\text{LikeCar}, \text{Color})$

$= 0.97 - 0.806$

$= 0.164$

VII. Split on Cost

	LikeCar	Yes	No
Cost	High	1	0
	Low	2	2

$E(\text{LikeCar}, \text{Cost}) = P(\text{High}) \cdot E(1,0) + P(\text{Low}) \cdot E(2,2)$

$= \frac{1}{5} \cdot E(1,0) + \frac{4}{5} \cdot E(2,2)$

$= \frac{1}{5}(0.46) + \frac{4}{5}(1.05) = 0.933$

$E(1,0) = E(1/5, 0/5)$

$= -\left(\frac{1}{5}\log_2\frac{1}{5}\right) = 0.46$

$IG(\text{LikeCar}, \text{Color}) = E(\text{LikeCar}) - E(\text{LikeCar}, \text{Color})$

$= 0.97 - 0.933$

$= 0.037$

$E(2,2) = E(2/5, 2/5)$

$= -\left(\frac{2}{5}\log_2\frac{2}{5}\right) - \left(\frac{2}{5}\log_2\frac{2}{5}\right)$

$= 1.05$

from V, VI, VII,

$IG(\text{LikeCar}, \text{Make}) = 0.164$
$IG(\text{LikeCar}, \text{Colour}) = 0.164$
$IG(\text{LikeCar}, \text{Cost}) = 0.037$

$\therefore$ we pick the first attribute with the highest information gain,
so we split on make

Decision Tree so far...

Type

Sedan | SUV

Make | No

Toyota | Ford

Toyota	Ford
Blue, High, ✓	Blue, Low, ✓
White, Low, X	White, Low, ✓
White, Low, X	

Entropy (Like Car) = E(0,2)

$= -\left(\frac{2}{2}\log_2\frac{2}{2}\right)$

$= \underline{\underline{0}}$

Entropy (Like Car) $= E(2,1)$

$= E(2/3, 1/3)$

$= -\left(\frac{2}{3}\log_2\frac{2}{3}\right) - \left(\frac{1}{3}\log_2\frac{1}{3}\right)$

$= \underline{\underline{0.92}}$

Decision Tree so far....

Type

Sedan | SUV

Make | No

Toyota | Ford

Blue, High, ✓

White, Low, X

White, Low, X

Ford: Yes

VIII. Split on Color

	LikeCar	Yes	No
Color	Blue	1	0
	White	0	2

$E(1,0) = E(1/3, 0)$

$= -\left(\frac{1}{3}\log_2 \frac{1}{3}\right)$

$= 0.528$

$E(0,2) = E(0, 2/3)$

$= -\left(\frac{2}{3}\log_2 \frac{2}{3}\right)$

$= 0.39$

$E(\text{LikeCar, Color}) = P(\text{Blue}) \cdot E(1,0) + P(\text{White}) \cdot E(0,2)$

$= \frac{1}{3} \cdot E(1,0) + \frac{2}{3} \cdot E(0,2)$

$= \frac{1}{3}(0.528) + \frac{2}{3}(0.39)$

$= 0.43$

$IG(\text{LikeCar, Color}) = E(\text{LikeCar}) - E(\text{LikeCar, Color})$

$= 0.92 - 0.43 = 0.49$

IX. Split on Cost

	LikeCar	Yes	No
Cost	High	1	0
	Low	0	2

$E(\text{LikeCar, Cost}) = P(\text{High}) \cdot E(1,0) + P(\text{Low}) \cdot E(0,2)$

$= \frac{1}{3} \cdot E(1,0) + \frac{2}{3} \cdot E(0,2)$

$= 0.43$

$IG(\text{LikeCar, Cost}) = E(\text{LikeCar}) - E(\text{LikeCar, Cost})$

$= 0.92 - 0.43 = 0.49$

from VIII, IX;

$IG(\text{LikeCar, Color}) = 0.49$

$IG(\text{LikeCar, Cost}) = 0.49$

∴ we pick the first attribute with the highest information gain, so we split on color.

Decision Tree...

2-level only

Blue	White
High, ✓	Low, ✗
	Low, ✗

Entropy = 0,
hence leaf nodes