CMSC 409: Artificial Intelligence
Project No. 1

Heman Baral
Jedidiah Pottle
James Stallings

CMSC 409: Artificial Intelligence
Project No. 1
Due Thursday, September 1217, 2019, noon

Student certification:
Team member 1:
Print Name: Heman Baral
Date: 17 Sep 2019

I have contributed by doing the following:

For dataset A,
Verified Confusion matrix data

For dataset B,
Verified Confusion matrix data

For dataset C,
Found the formula for the activation line
Constructed the model of the neuron (graphically and mathematically)
Verified Confusion matrix data

For Part B,
with Jedidiah Pottle and James Stallings, created the Truth Table and answered all of part B questions.

Signed : heman ___________________ *(you can sign/scan or use e-signature)*

CMSC 409: Artificial Intelligence
Project No. 1
Due Thursday, September 1217, 2019, noon

Student certification:
Team member 2:
Print Name: Jedidiah Pottle
Date: 17 Sep 2019

I have contributed by doing the following:

For dataset A,
With Jim Stallings, constructed the model of the neuron
Computed Confusion matrix and percentages

For dataset B,
Computed Confusion matrix and percentages

For dataset C,
Computed Confusion matrix and percentages

Wrote the bulk of the Conclusions paragraph.

with Heman Baral and James Stallings, created the Truth Table and answered all of part B questions.

Signed : Jedidiah J Pottle *(you can sign/scan or use e-signature)*

CMSC 409: Artificial Intelligence
Project No. 1
Due Thursday, September 1217, 2019, noon

Student certification: Team member 3:
Print Name: ***James Stallings***
Date: 17 Sep 2019
I have contributed by doing the following:

Created a private project GitHub at https://github.com/coffee247/AI-Homework
Loaded the datasets into google sheets
Created charts that accurately plot the data points in each dataset

For dataset A,
Found the formula for the activation line
With Jedidiah Pottle, constructed the model of the neuron
Created the chart & drew the line
Verified Confusion matrix data

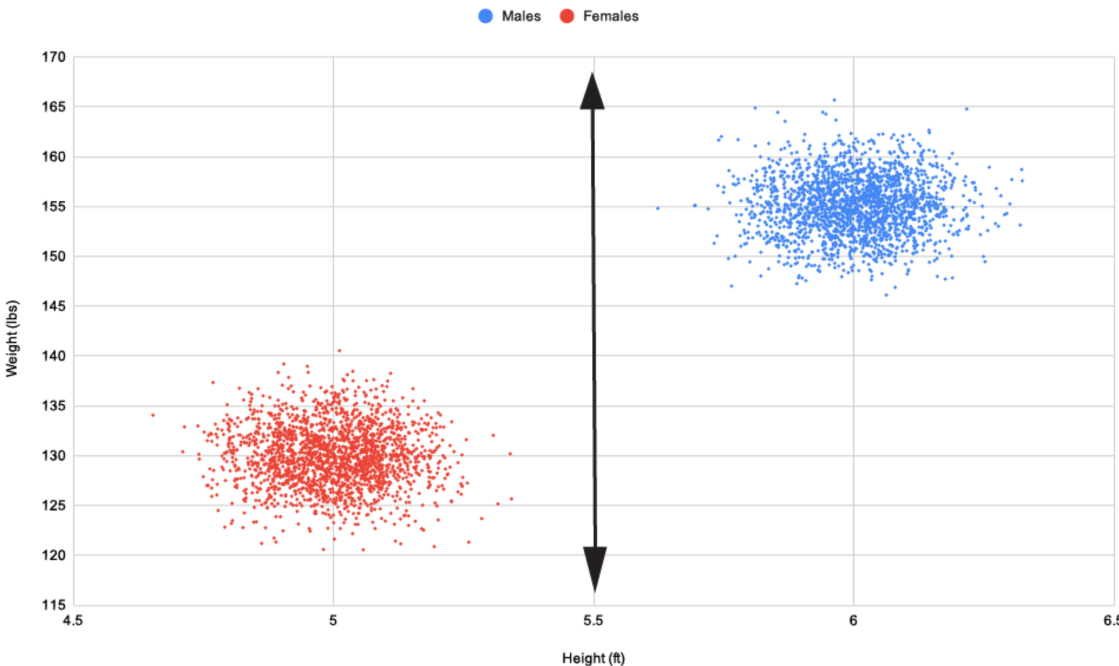
For dataset B,
Found the formula for the activation line
Constructed the model of the neuron (graphically and mathematically)
Created the chart & drew the line
Verified Confusion matrix data

For dataset C,
Constructed the model of the neuron (graphically and mathematically)
Created the chart & drew the line
Verified Confusion matrix data

For Part B,
with Jedidiah Pottle and Heman Baral, created the Truth Table and answered all of part B questions.

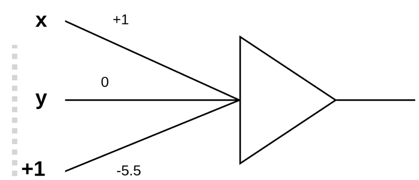
Signed : James M. Stallings *(you can sign/scan or use e-signature)*

DataSetA analysis



$$net = \sum_{i=1}^{n} \omega_i x_i + \omega_{n+1};$$

$$0 = \begin{Bmatrix} 1, if net \geq 0 \\ 0, if net < 0 \end{Bmatrix}$$



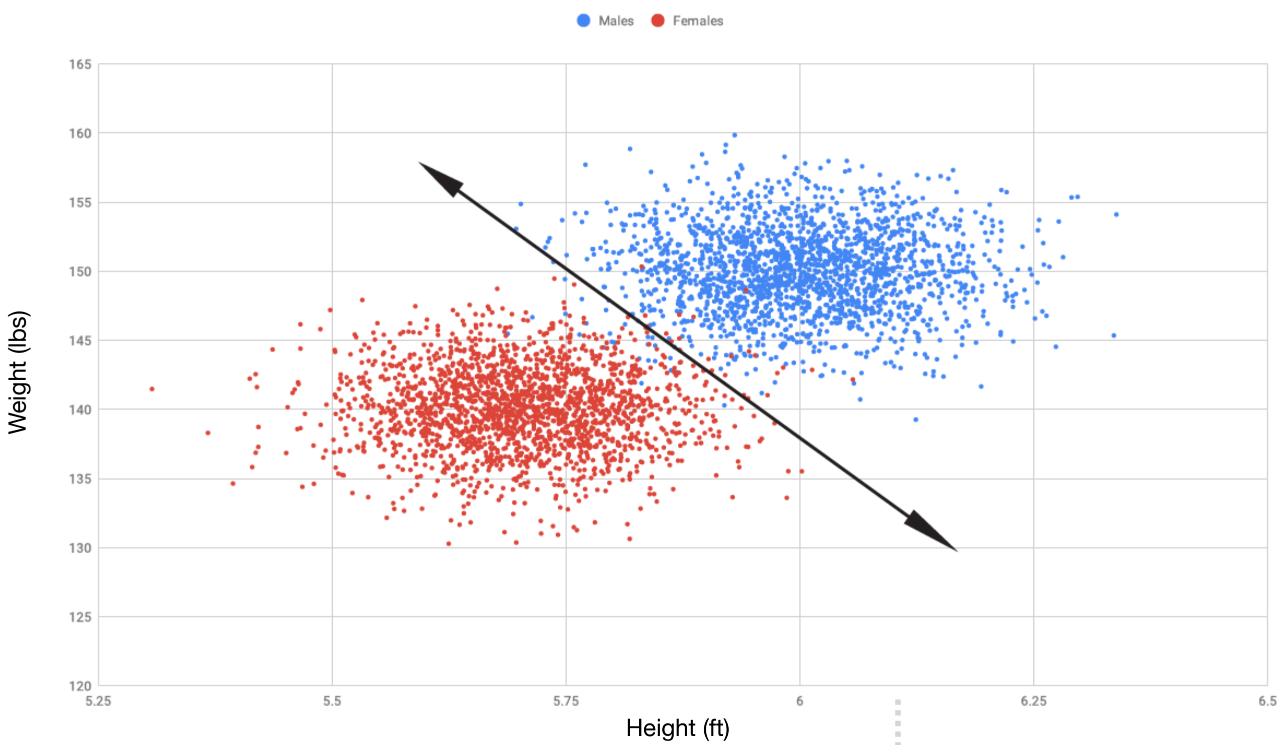
The equation of the linear separator is: $x = 5.5$

The definition of the neuron is: x - 5.5 ≥ 0

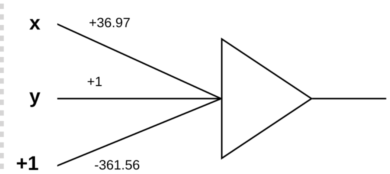
Male = Positive … Female = Negative
The error rate is zero (all data points are properly classified.)
There are no false positives FP=0. There were 2000 True Positives TP=2000.
There are no false negatives FN=0. There were 2000 True Negatives TN=2000.

False Negative Rate = FN/(FN+TP) = 0%
False Positive Rate = FP/(FP+TN) = 0%
True Positive Rate = TP/(TP+FN) = 100%
True Negative Rate = TN/(TN+FP) = 100%

DataSetB analysis



$$net = \sum_{i=1}^{n} \omega_i x_i + \omega_{n+1}; \qquad 0 = \begin{Bmatrix} 1, if net \geq 0 \\ 0, if net < 0 \end{Bmatrix}$$



The equation of the linear separator is: $y = (-36.97(x)) + 361.56$

The equation of the neuron is 36.97(x) +y - 361.56 $\geq$ 0

Male = Positive … Female = Negative

There are 22 false positives FP=22.

There are 27 false negatives FN=27.

There were 1973 True Positives TP=1973.

There were 1978 True Negatives TN=1978.

False Negative Rate = FN/(FN+TP) = (27/(27+1973))*100 = 1.35%

False Positive Rate = FP/(FP+TN) = (22/(22+1978))*100 = 1.10%

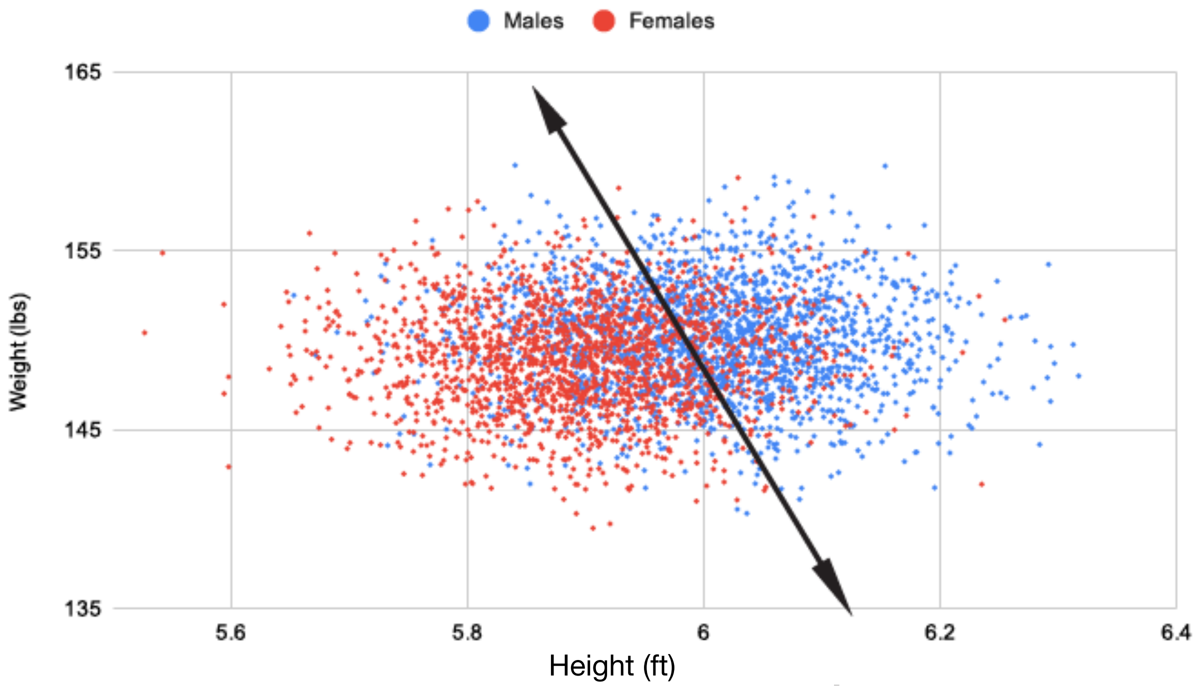
True Positive Rate = TP/(TP+FN) = (1973/(1973+27))*100 = 98.65%

True Negative Rate = TN/(TN+FP) = (1978/(1978+22))*100 = 98.90%

The total error is ((FP+FN)/(TP+TN)) = ((27+22)/(1973+1978))*100 = 1.24%

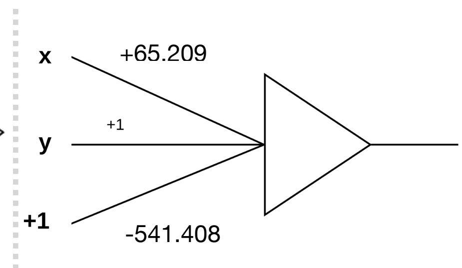
Pottle, Jedidiah Baral, Heman Stallings, James

DataSetC analysis



$$net = \sum_{i=1}^{n} \omega_i x_i + \omega_{n+1};$$

$$0 = \begin{Bmatrix} 1, if net \geq 0 \\ 0, if net < 0 \end{Bmatrix}$$



The equation of the linear separator is: $y = (-65.209(x)) + 541.408$

The equation of the neuron is 65.209(x) +y - 541.408 $\geq$ 0

Male = Positive … Female = Negative

There are 246 false positives FP=246.

There are 988 false negatives FN=988.

There were 1012 True Positives TP=1012.

There were 1571 True Negatives TN=1571.

False Negative Rate = FN/(FN+TP) = (988/(988+1012))*100 = 49.40%

False Positive Rate = FP/(FP+TN) = (246/(246+1571))*100 = 13.54%

True Positive Rate = TP/(TP+FN) = (1012/(1012+988))*100 = 50.60%

True Negative Rate = TN/(TN+FP) = (1571/(1571+246))*100 = 86.46%

The total error is ((FP+FN)/(TP+TN)) = ((988+246)/(1012+1571))*100 = 47.77%

Pottle, Jedidiah Baral, Heman Stallings, James

DataSet Result Comparisons

Data-sets A, B, and C, are three hypothetical data-sets intended for use as training data when building an AI model that classifies new data as male or female.

When plotted in a two-dimensional Cartesian plane, the three data-sets exhibit largely different clustering behaviors relative to the known sexes of the data-set values.

With data set A's Male and Female clusters, the two clusters are farther apart than in data set B or in data set C. This allows Males and Females to be easily separated using either weight or height as only one characteristic is needed to discern Males from Females, and vice versa.

In data set B, the Male and the Female clusters are much closer together than in data set A, causing a slight overlap between the two clusters. This makes it a little more difficult to separate Males from Females, and vice versa.

For data set C, the Male and the Female clusters are on top of each other, making it very difficult to properly differentiate Males from Females, and vice versa.

The definition of data set A's neuron had the better accuracy (100%) of all the data sets as there was no overlap of Males and Females. Data set B's neuron gave an accuracy close to that of data set A (98.76%) since there was very little overlap between the Male and Female clusters. Data set C's neuron definition showed the worst accuracy (52.23%) due to both clusters nearly occupying the same space.

From the neuron model of data set A to data set B to data set C, the weights for the height increment into the positives and the weights for the bias decrement into the negatives. This shows that each data set requires a greater emphasis on the importance of the inputs. Primarily, the height and the bias are seen to have a greater presence than weight. In data set A, weight is not even a factor.

Part B

Truth Table					
A	B	C	B + AB	Inequalities	out
0	0	0	0	0<T, T= 1.5	0 < 1.5
0	0	1	0	wc < T	-2 < 1.5
0	1	0	1	wb >= T	3 >= 1.5
0	1	1	0	wb + wc < T	1 < 1.5
1	0	0	0	wa < T	-1 < 1.5
1	0	1	0	wa + wc < T	-3 < 1.5
1	1	0	1	wa + wb >= T	2 >= 1.5
1	1	1	0	wa + wb + wc < T	0 < 1.5

The functionality of the neuron is: **B or AB**

The range of possible values for the threshold (T) that will produce a 1 as output: **(1,2]**

Pottle, Jedidiah Baral, Heman Stallings, James