Homework Assignment # 5

Solution 1:

a) Given dataset can be represented as following in tabular form:

Datapoint	$X_{i,0}$	$X_{i,1}$	Y_{true}	$Y_{predict}$
X_1	0	0	0	?
X_2	0	1	1	?
X_3	1	0	1	?
X_4	1	1	0	?

We need to find the $Y_{predict}$values for each point X_i in dataset using Naive Bayes classification.

The condition for classifying point as positive is as follows:

$P(1).P(x|1) \succcurlyeq P(0).P(x|0)$

Since features will be considered independent due to naive assumption we can write the above rule as following:

$Y_{pred} = 1$if $P(1).\prod p(x_j|1) \succcurlyeq p(0).\prod p(x_j|0)$

Let us write the 0 and 1 class conditional probabilities ($p(x|Class)$) for both features x_1 *and* $x_2$in D to work on the above rule.

$p(x_1 = 1|0) = \frac{1}{2}, p(x_2 = 1|0) = \frac{1}{2}$
$p(x_1 = 0|0) = \frac{1}{2}, p(x_2 = 0|0) = \frac{1}{2}$
$p(x_1 = 1|1) = \frac{1}{2}, p(x_2 = 0|1) = \frac{1}{2}$
$p(x_1 = 0|1) = \frac{1}{2}, p(x_2 = 0|1) = \frac{1}{2}$

As we can see above all the possible combinations of feature given class pairs results in the probability of 0.5

We also know the prior probabilities of classes as p(0) = p(1) = $\frac{1}{2}$

Now we can apply the rule stated earlier for all points, so lets work.

Datapoint	$X_{i,0}$	$X_{i,1}$	Y_{true}	$P(1).\prod p(x_j|1)$	$p(0).\prod p(x_j|0)$	Y_{pred}
X_1	0	0	0	$\frac{1}{2} * \frac{1}{2} = \frac{1}{4}$	$\frac{1}{2} * \frac{1}{2} = \frac{1}{4}$	1
X_2	0	1	1	$\frac{1}{2} * \frac{1}{2} = \frac{1}{4}$	$\frac{1}{2} * \frac{1}{2} = \frac{1}{4}$	1
X_3	1	0	1	$\frac{1}{2} * \frac{1}{2} = \frac{1}{4}$	$\frac{1}{2} * \frac{1}{2} = \frac{1}{4}$	1
X_4	1	1	0	$\frac{1}{2} * \frac{1}{2} = \frac{1}{4}$	$\frac{1}{2} * \frac{1}{2} = \frac{1}{4}$	1

Since for all points $P(1).\prod p(x_j|1) = p(0).\prod p(x_j|0)$ we will assign positive class to all the points.

Now we calculate the accuracy of our predictions according to following formula:

$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$

In this prediction we have TP=2, FN=2, TN=0, FP=0

Substituting we get, $Accuracy = \frac{2}{4} or$ 50%

b) We will follow the exact same process as part a) here:

Given data in tabular form:

Datapoint	1	$X_{i,0}$	$X_{i,1}$	$X_{i,0}$*$X_{i,1}$	$X_{i,0}^2$	$X_{i,1}^2$	Y_{true}	$Y_{predict}$
X_1	1	0	0	0	0	0	0	?
X_2	1	0	1	0	0	1	1	?
X_3	1	1	0	0	1	0	1	?
X_4	1	1	1	1	1	1	0	?

Then we write the class conditional probabilities of all the features:

Features	p(0|0)	p(0|1)	p(1|0)	p(1|1)
1	0	0	1	1
$X_{i,0}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$X_{i,1}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$X_{i,0}$*$X_{i,1}$	$\frac{1}{2}$	1	$\frac{1}{2}$	0
$X^2_{i,0}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$X^2_{i,1}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

Now we can calculate the probabilties required for Naive Bayes classification:

Datapoint	Y_{true}	$P(1).\prod p(x_j\|1)$	$p(0).\prod p(x_j\|0)$	Y_{pred}
X_1	0	$\frac{1}{2}*1*\frac{1}{2}*\frac{1}{2}*1*\frac{1}{2}*\frac{1}{2}=\frac{1}{2^5}$	$\frac{1}{2}*1*\frac{1}{2}*\frac{1}{2}*\frac{1}{2}*\frac{1}{2}*\frac{1}{2}=\frac{1}{2^6}$	1
X_2	1	$\frac{1}{2}*1*\frac{1}{2}*\frac{1}{2}*1*\frac{1}{2}*\frac{1}{2}=\frac{1}{2^5}$	$\frac{1}{2}*1*\frac{1}{2}*\frac{1}{2}*\frac{1}{2}*\frac{1}{2}*\frac{1}{2}=\frac{1}{2^6}$	1
X_3	1	$\frac{1}{2}*1*\frac{1}{2}*\frac{1}{2}*1*\frac{1}{2}*\frac{1}{2}=\frac{1}{2^5}$	$\frac{1}{2}*1*\frac{1}{2}*\frac{1}{2}*\frac{1}{2}*\frac{1}{2}*\frac{1}{2}=\frac{1}{2^6}$	1
X_4	0	$\frac{1}{2}*1*\frac{1}{2}*\frac{1}{2}*0*\frac{1}{2}*\frac{1}{2}=0$	$\frac{1}{2}*1*\frac{1}{2}*\frac{1}{2}*\frac{1}{2}*\frac{1}{2}*\frac{1}{2}=\frac{1}{2^6}$	0

$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$

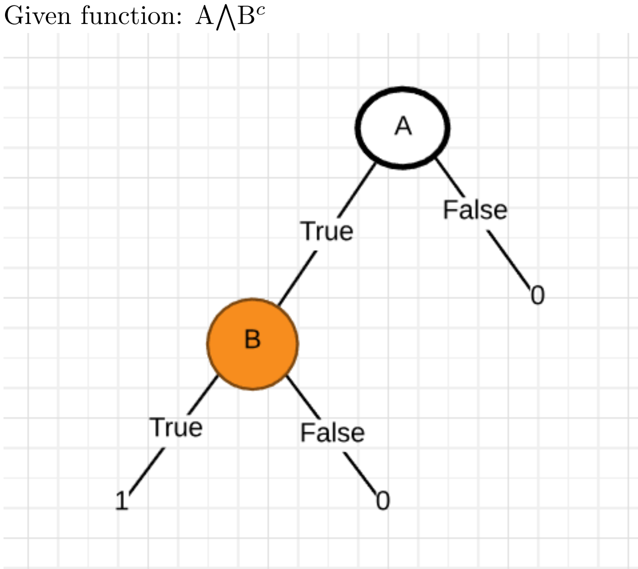
In this prediction we have TP=2, FN=1, TN=1, FP=0

Substituting we get, $Accuracy = \frac{3}{4} or$ 75%

Solution 2:

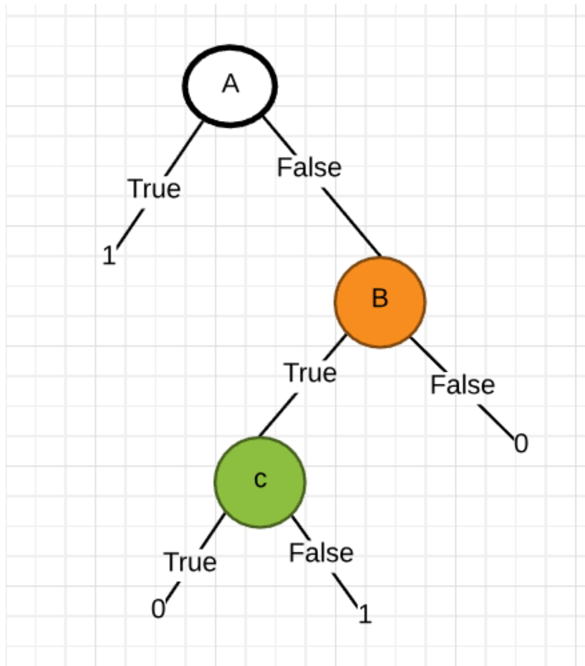
a)

Given function: A$\bigwedge$B^c

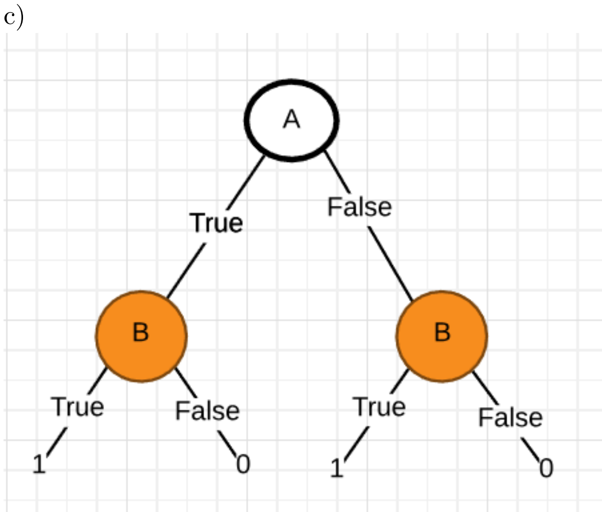


b)

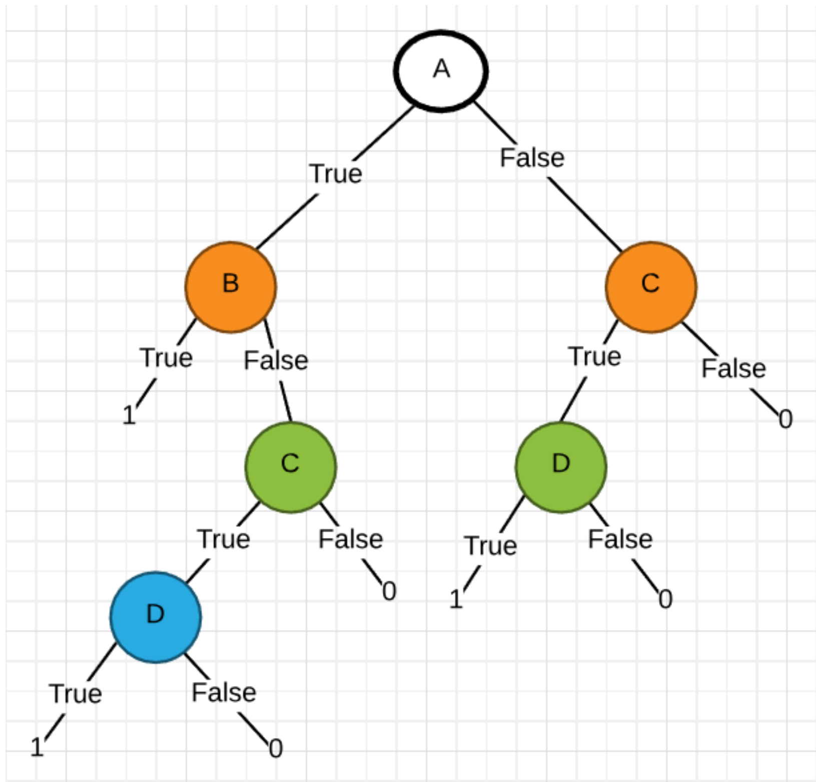
Given Function: A $\bigvee$ (B$\bigwedge$C^c)


A
True
False
1
B
True
False
0
c
True
False
0
1

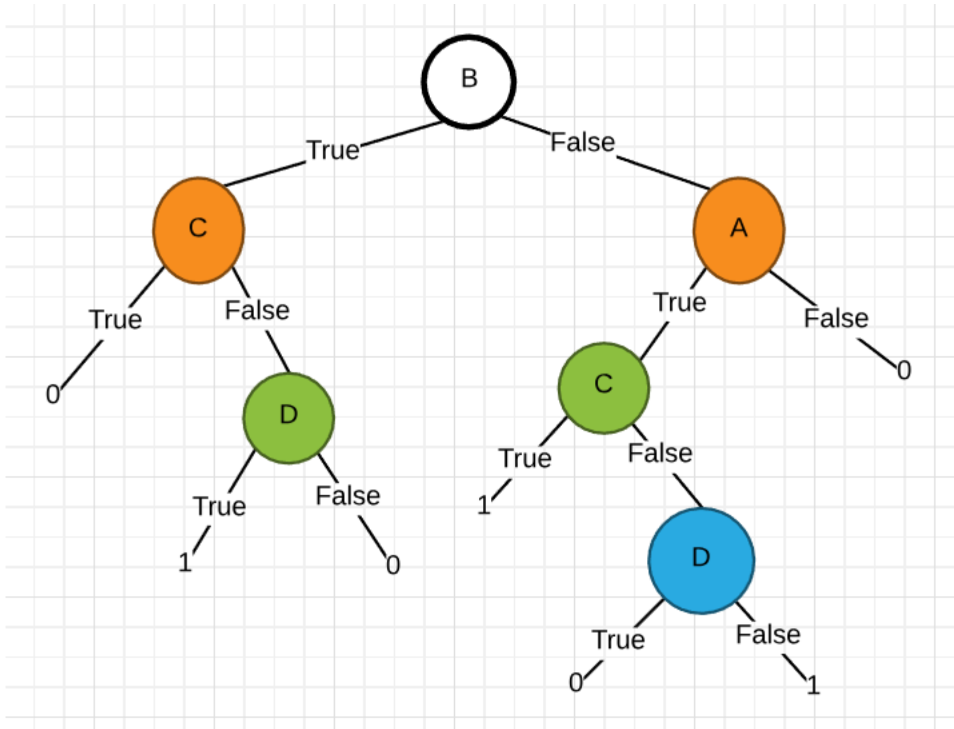
c)


A
True
False
B
B
True
False
True
False
1
0
1
0

d)


A
True
False
B
C
True
False
True
False
1
0
C
D
True
False
True
False
0
1
0
D
True
False
1
0

e)



Solution 3:

a)

We will start by calculating the Entropy at parent node denoted by E(p)

$E(p) = -\frac{3}{4}log\frac{3}{4} - \frac{1}{4}log\frac{1}{4} = 0.8112$

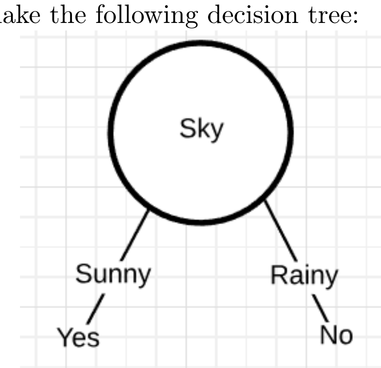
Now we will calculate the Information Gain denoted by GAIN$_{columnname}$

First we look at column Sky and calculate the Entropies for each split say E(Sunny) and E(Rainy)

$E(Sunny) = -\frac{3}{3}log\frac{3}{3} - 0 = 0, E(Rainy) = -\frac{1}{1}log\frac{1}{1} - 0 = 0$

$GAIN_{sky} = E(p) - \frac{3}{4}E(sunny) - \frac{1}{4}E(Rainy) = E(p) = 0.8112$

Since this is the maximum gain we will skip looking at further columns and make the following decision tree:



b)

First we will calculate the changed entropy at parent node:

$E(p) = -\frac{3}{5}log\frac{3}{5} - \frac{2}{5}log\frac{2}{5} = 0.9709$

We can start by calculating Information Gain for each column:

Sky: $E(Sunny) = -\frac{3}{4}log\frac{3}{4} - \frac{1}{4}log\frac{1}{4} = 0.8112, E(Rainy) = -\frac{1}{1}log\frac{1}{1} - 0 = 0$

$GAIN_{sky} = E(p) - \frac{4}{5}E(sunny) - \frac{1}{5}E(Rainy) = 0.9709 - 0.6489 = 0.322$

Temperature: $E(Warm) = -\frac{3}{4}log\frac{3}{4} - \frac{1}{4}log\frac{1}{4} = 0.8112, E(Cold) = -\frac{1}{1}log\frac{1}{1} - 0 = 0$

$GAIN_{temperature} = E(p) - \frac{4}{5}E(Warm) - \frac{1}{5}E(Cold) = 0.9709 - 0.6489 = 0.322$

Humidity: $E(Normal) = -\frac{1}{2}log\frac{1}{2} - \frac{1}{2}log\frac{1}{2} = 1, E(High) = -\frac{2}{3}log\frac{2}{3} - \frac{1}{3}log\frac{1}{3} = 0.9182$

$GAIN_{humidity} = E(p) - \frac{2}{5}E(Normal) - \frac{3}{5}E(High) = 0.0199$

Wind: $E(Strong) = -\frac{3}{4}log\frac{3}{4} - \frac{1}{4}log\frac{1}{4} = 0.8112, E(Weak) = 0$

$GAIN_{wind} = E(p) - \frac{4}{5}E(Strong) - \frac{1}{5}E(Weak) = 0.322$

Water: $E(Warm) = -\frac{2}{4}log\frac{2}{4} - \frac{2}{4}log\frac{2}{4} = 1, E(Cool) = 0$

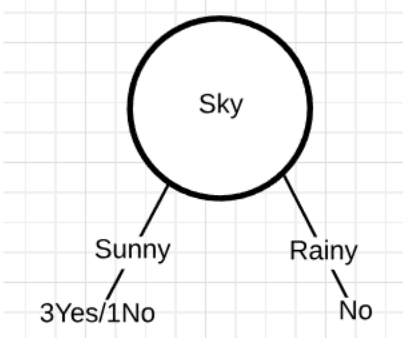
$GAIN_{water} = E(p) - \frac{4}{5}E(Warm) - \frac{1}{5}E(Cool) = 0.1709$

Forecast: $E(Same) = -\frac{2}{3}log\frac{2}{3} - \frac{1}{3}log\frac{1}{3} = 0.9182, E(Change) = -\frac{1}{2}log\frac{1}{2} - \frac{1}{2}log\frac{1}{2} = 1$

$GAIN_{water} = E(p) - \frac{3}{5}E(Same) - \frac{2}{5}E(Change) = 0.0199$

From above calculations we can conclude that splitting by 3 columns Sky, Temperature and Wind will result in GAIN of 0.322.

Therefore we split by Sky and get the following decision tree:



We further need to split the left node as it does not contain similar class so we repeat GAIN calculations:

$E(p) = -\frac{3}{4}log\frac{3}{4} - \frac{1}{4}log\frac{1}{4} = 0.8112$

Temperature: $E(Warm) = -\frac{3}{4}log\frac{3}{4} - \frac{1}{4}log\frac{1}{4} = 0.8112$

$GAIN_{temperature} = E(p) - \frac{4}{4}E(Warm) = 0$

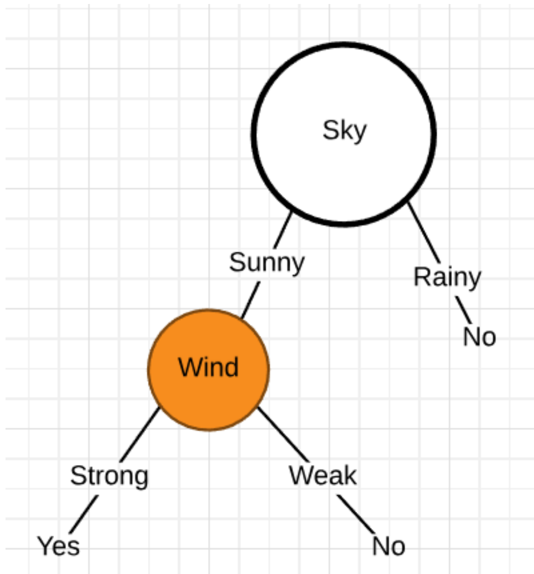
Humidity: $E(Normal) = -\frac{1}{2}log\frac{1}{2} - \frac{1}{2}log\frac{1}{2} = 1, E(High) = 0$

$GAIN_{humidity} = E(p) - \frac{2}{4}E(Normal) = 0.3112$

Wind: $E(Strong) = 0, E(Weak) = 0$

$GAIN_{wind} = E(p) - 0 = 0.8112$

Since this is the maximum attainable gain we will stop here and Split the tree on Wind:



Solution 4:

We will consider the following datasets for hidden neurons plot vs classfication accuracy and epochs vs accuracy plots:

NOTE:

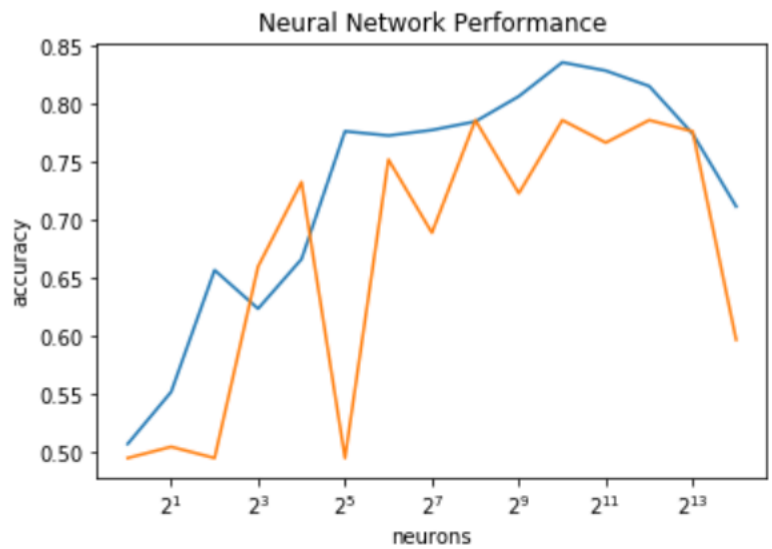
a) BLUE line is training set and YELLOW line is test set

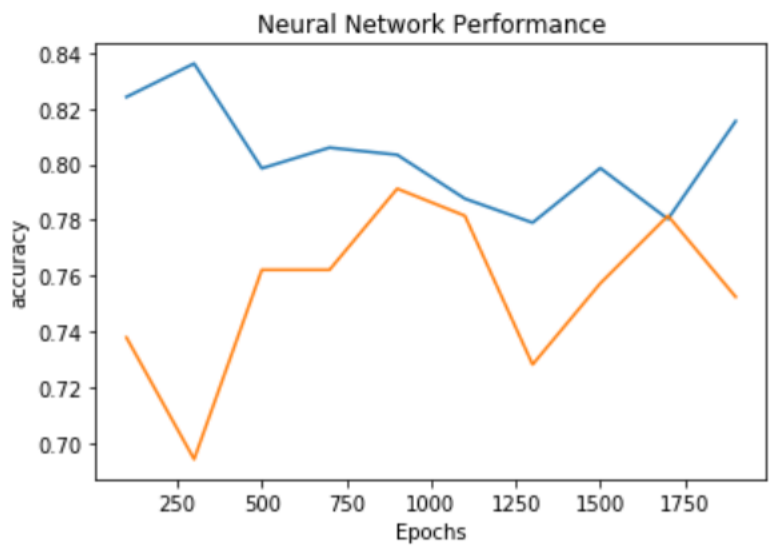
b) For variations in epochs we plot accuracy but can see error trends as error will be (1-accuracy)

c) Training accuracies are 5-Fold cross validated

Tool used for entire analysis is sklearn.neural_network's object MLPClassifier and implementation can be found in file problem4.py

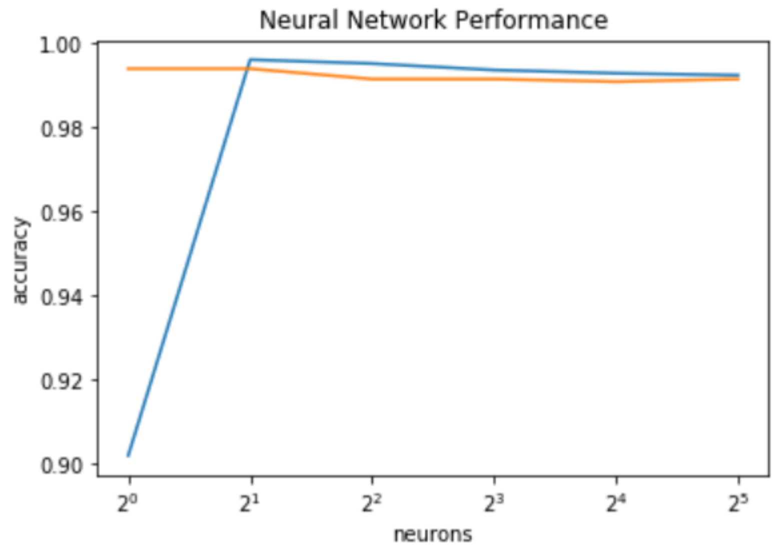
1. concrete

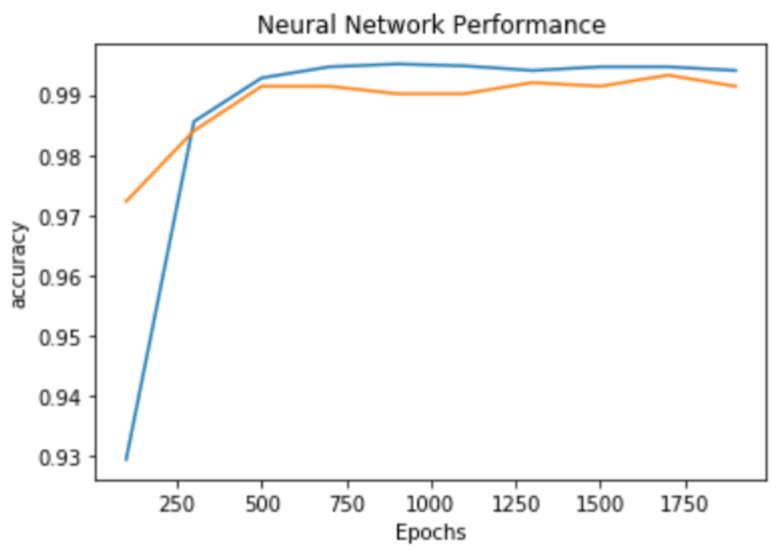




Classification accuracy increases with both neurons and epochs. (Sudden dip in accuracy at 2^{14}neurons might be due to overfitting)

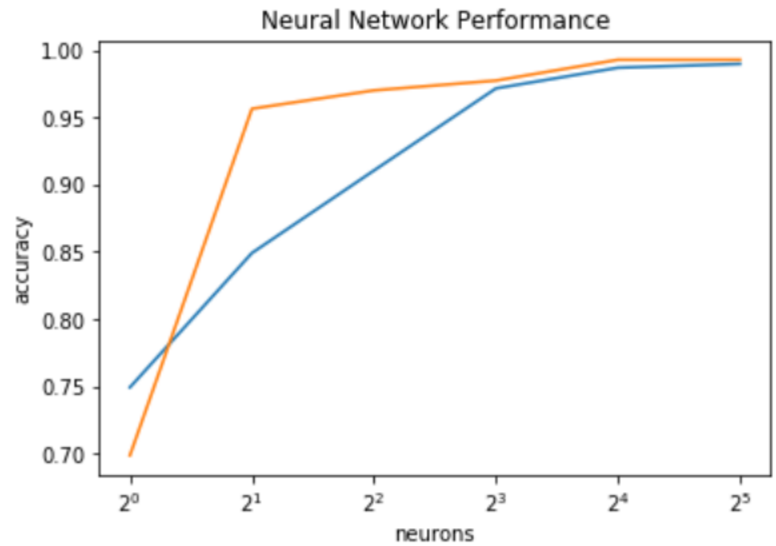
2. mushroom

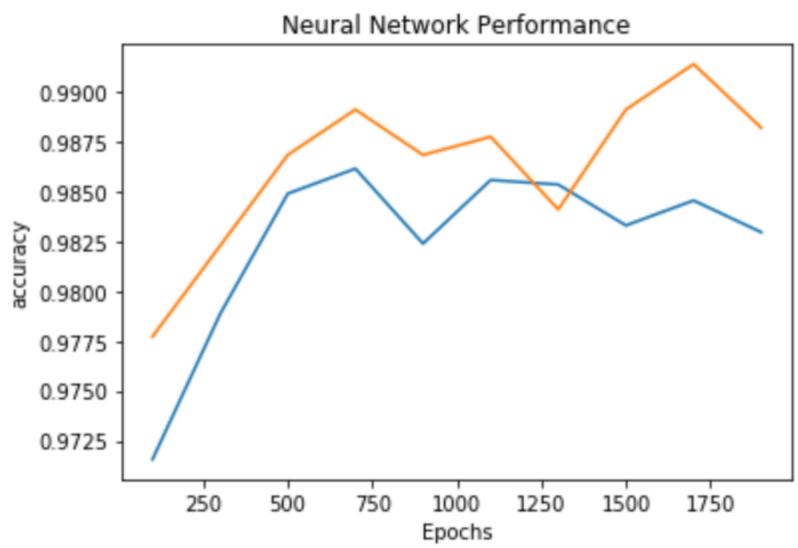




Classification accuracy increases with both neurons and epochs consistently.

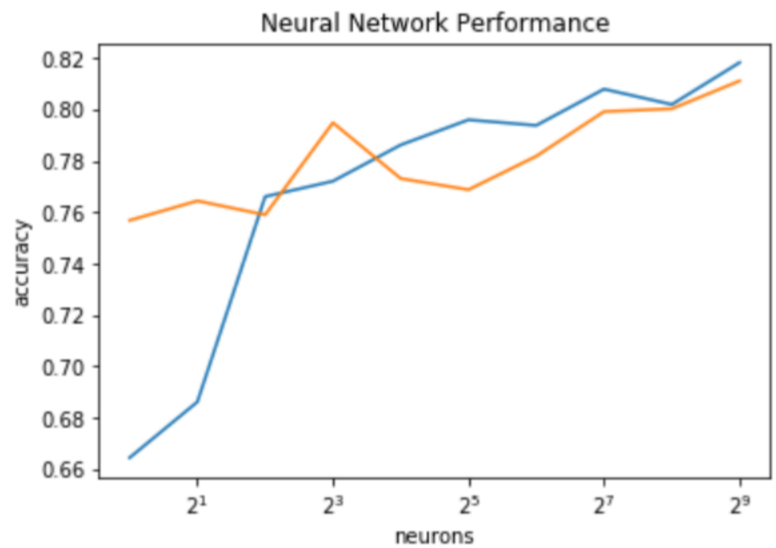
3. pendigit

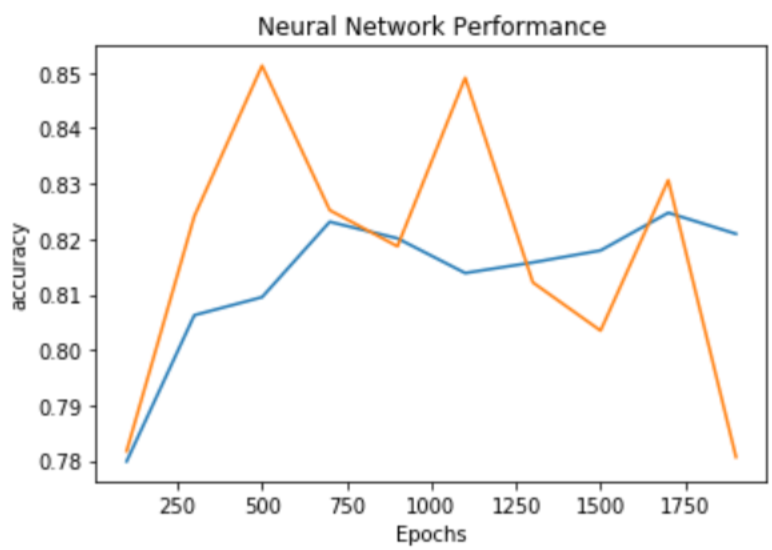




Classification accuracy increases with neurons and epochs consistently (Minor fluctuations in variation with epochs)

4. spambase

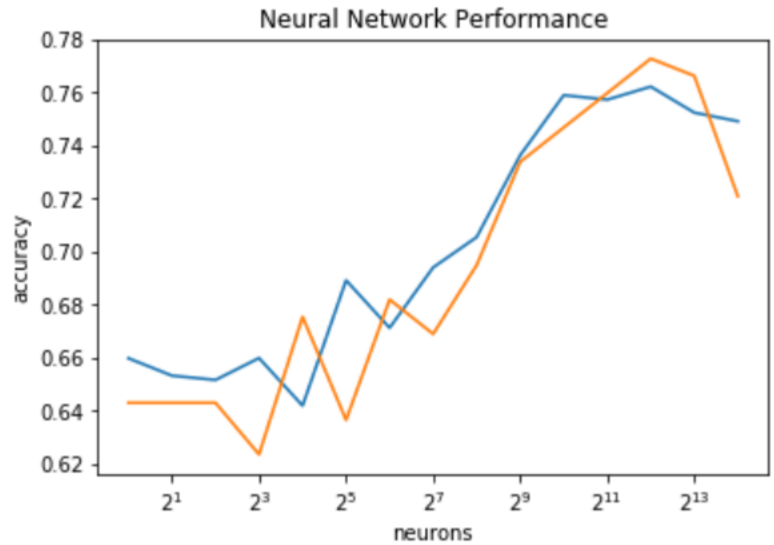


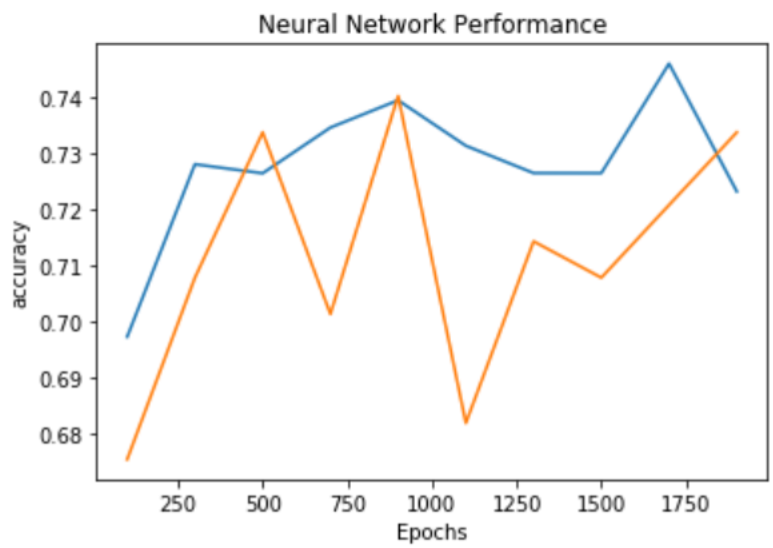


Classification accuracy increases with neurons. For epochs training accuracy increases consistently although

some fluctuations in test accuracy.

5. pima





Classification accuracy increases with neurons. For epochs training accuracy increases consistently although

some fluctuations in test accuracy.

Finally we can conclude the claim that with increasing number of neurons the accuracy of a neural network is destined to increase.

Although, a few cases might result in overfitting.