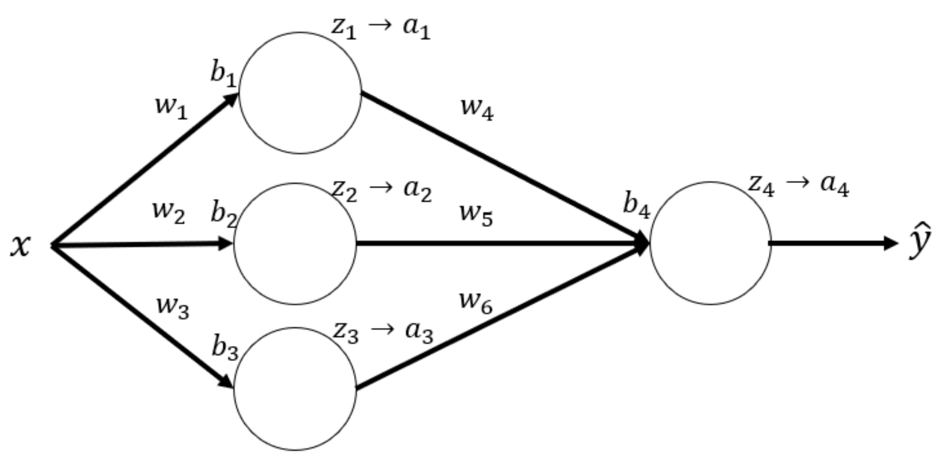
**Homework 5** **Name: ____________________**
**ECE/CS 498 DS Spring 2020** **NetID: ____________________**
**Issued: 04/27/20**
**Due: 05/04/20 23:59:59 on Compass**

***What to submit:***

- Jupyter notebook for coding part of problem 2
- A PDF report with the calculations of problem 1 and the results from problem 2.

## Problem 1

Consider the neural network given below. Assume that all the neurons use the sigmoid activation function.




1. **(3 points)** Write down the expressions for $z_1, z_4, a_1$ and $a_4$. Use them to express the output of the neural network $\hat{y}$ as a function of an input $x_1$, weights $w_1, w_2, w_3, w_4, w_5, w_6$, and biases $b_1, b_2, b_3, b_4$

$$z_1 = w_1 x_1 + b_1$$
$$a_1 = \sigma(z_1)$$
$$z_4 = \sigma(z_1)w_4 + \sigma(z_2)w_5 + \sigma(z_3)w_6$$
$$a_2 = \sigma(z_2)$$

$$\hat{y} = a_2$$
$$= \sigma(\sigma(w_1 x_1 + b_1)w_4 + \sigma(w_2 x_1 + b_2)w_5 + \sigma(w_3 x_1 + b_3)w_6 + b_4)$$

**Name:** ____________________
**NetID:** ____________________

2. **(5 points; 1 for each)** In the following questions, we use mean squared error as the loss function $L$. Consider a single sample $x_i$ , with predicted label $\hat{y}_i$ and true label $y_i$. Write down the equation for $L_i$ and for the gradients $\frac{\partial L_i}{\partial w_1}$, $\frac{\partial L_i}{\partial w_4}$, $\frac{\partial L_i}{\partial b_1}$, and $\frac{\partial L_i}{\partial b_4}$. (You may use $\hat{y}_i, y_i, z_{1i}, z_{2i}, z_{3i}, z_{4i}, w_1, w_2, w_3, w_4, w_5, w_6$, $b_1, b_2, b_3, b_4$ and $x_i$ in your expressions.)

$$L_i = (\hat{y}_i - y_i)^2$$

$$\frac{\partial L_i}{\partial w_4} = \frac{\partial L_i}{\partial a_4}\frac{\partial a_4}{\partial z_{4i}}\frac{\partial z_{4i}}{\partial w_4}$$
$$= 2(\hat{y}_i - y_i)\sigma(z_{4i})\sigma(-z_{4i})\sigma(z_{1i})$$

$$\frac{\partial L_i}{\partial b_4} = \frac{\partial L_i}{\partial a_4}\frac{\partial a_4}{\partial z_{4i}}\frac{\partial z_{4i}}{\partial b_4}$$
$$= 2(\hat{y}_i - y_i)\sigma(z_{2i})\sigma(-z_{2i})$$

$$\frac{\partial L_i}{\partial w_1} = \frac{\partial L_i}{\partial a_4}\frac{\partial a_4}{\partial z_{4i}}\frac{\partial z_{4i}}{\partial a_1}\frac{\partial a_1}{\partial z_{1i}}\frac{\partial z_{1i}}{\partial w_1}$$
$$= 2(\hat{y}_i - y_i)\sigma(z_{4i})\sigma(-z_{4i})w_4\sigma(z_{1i})\sigma(-z_{1i})x_i$$

$$\frac{\partial L_i}{\partial b_1} = \frac{\partial L_i}{\partial a_4}\frac{\partial a_4}{\partial z_{4i}}\frac{\partial z_{4i}}{\partial a_1}\frac{\partial a_1}{\partial z_{1i}}\frac{\partial z_{1i}}{\partial b_1}$$
$$= 2(\hat{y}_i - y_i)\sigma(z_{2i})\sigma(-z_{2i})w_4\sigma(z_{1i})\sigma(-z_{1i})$$

Note: for sigmoid function, $\sigma(-z)$ is equal to $1 - \sigma(z)$

3. **(2 points)** Given a single training sample $x_1$, write the gradient descent equation for updating weight $w_1$.

$$w_1^{t+1} = w_1^t - \eta\frac{\partial L_1}{\partial w_1^t}$$
$$= w_1^t - \eta 2(\hat{y}_1 - y_1)\sigma(z_{41}^t)\sigma(-z_{41}^t)w_4^t\sigma(z_{11}^t)\sigma(-z_{11}^t)x_1$$

Where $\eta$ is the learning rate, and superscript $t$ denotes the value of the variable at the $t^{th}$ iteration of gradient descent.

4. **(2 points)** Given $n$ training samples $x_i$ for $i = 1,2,3, \dots, n$, write the gradient descent equation for updating weight $w_1$.

$$w_1^{t+1} = w_1^t - \frac{\eta}{n} \sum_{i=1}^{n} \frac{\partial L_i}{\partial w_1^t}$$

$$= w_1^t - \frac{\eta}{n} \sum_{i=1}^{n} 2(\hat{y}_i - y_i)\sigma(z_{4i}^t)\sigma(-z_{4i}^t)w_4^t\sigma(z_{1i}^t)\sigma(-z_{1i}^t)x_i$$

Where $\eta$ is the learning rate, and superscript $t$ denotes the value of the variable at the $t^{th}$ iteration of gradient descent.

## Problem 2

Machine Learning has seen wide applications, one of which includes physics and astronomy. You might have heard about the recent achievement of imaging a black hole. The method used in that work involved machine learning. In this homework, you'll explore the use of SVM, Decision Trees and Random Forests classifiers in another problem based in astronomy.

***Data:*** *HTRU_2.csv*

HTRU2 (High Time Resolution Universe Survey) Dataset. It describes a sample of pulsar candidates collected during the survey. Pulsar is a rate type of Neutron star. See the link for details about the study and description of the features in the dataset: [https://archive.ics.uci.edu/ml/datasets/HTRU2](https://archive.ics.uci.edu/ml/datasets/HTRU2)

The data has 9 columns - 8 of which are features and 'Class' is the label. Class=0: Not a pulsar, Class=1: Is pulsar.

**Homework 5** **Name: ____________________**
**ECE/CS 498 DS Spring 2020** **NetID: ____________________**
**Issued: 04/27/20**
**Due: 05/04/20 23:59:59 on Compass**

## *Problem:*

Use sklearn package for SVM, decision trees, random forests, accuracy, precision and recall. Classify the samples based on the 8 features using the following methods:

- Linear SVM – regularization parameter C in {0.1, 1, 10}
- Decision Trees – maximum depth in {3, 4, 6}
- Random Forests – number of trees in {5, 11, 13}, maximum depth = 5

1. **(9 points – 1 for each classifier)** For each of the above perform 5-fold cross-validation of the data with 80%-20% train test split. Report the mean accuracy, precision and recall on the test set. Report precision and recall for the positive class (pulsar). Also report the respective standard deviations. Provide the results in a table.

   Table 1: Performance of each classifier (classifier type along with regularization parameter). Mean and standard deviation for each metric are given. Standard deviation is given in parenthesis.

| | SVM (C) | | | Decision Tree (depth) | | | Random Forest (#trees) | | |
|---|---|---|---|---|---|---|---|---|---|
| | 0.1 | 1 | 10 | 3 | 4 | 6 | 5 | 11 | 13 |
| Accuracy | 97.8 (0.8) | 97.9 (0.7) | 97.9 (0.7) | 97.8 (0.6) | 97.9 (0.6) | 97.9 (0.6) | 97.8 (0.6) | 97.8 (0.7) | 97.8 (0.7) |
| Precision | 89.8 (11.2) | 89.9 (10.3) | 89.5 (10.7) | 84.8 (12.0) | 86.6 (11.8) | 86.4 (11.9) | 88.6 (10.7) | 89.0 (10.3) | 88.9 (11.2) |
| Recall | 78.3 (6.1) | 79.6 (6.1) | 79.7 (6.0) | 84.1 (4.3) | 82.4 (5.7) | 81.7 (7.6) | 78.6 (8.6) | 78.6 (8.2) | 78.2 (8.5) |

2. **(1 point)** Explain the high value of accuracy compared to precision and recall.
   Accuracy is much higher compared to precision and recall because of imbalance in the dataset. There are approximately 10 times as many samples for Class=0 (not pulsar) as compared to Class=1 (pulsar).

3. **(1 point)** Which classifier performs the best? Explain the reason why the picked classifier performed the best.
   All classifiers give approximately the same accuracy. Precision is higher for SVM and Random Forests but their recall is worse compared to Decision trees. We pick decision tree with max depth=3 as the best classifier since it gives approximately equal precision and recall and has a lower standard deviation for recall compared to the rest. Possible explanation for why the Decision tree with max depth=3 performs the best is (i) allows for non-linear separation of the feature space, and (ii) limited depth prevents overfitting.

4. **(Optional)** Visualize a decision tree (depth=3) from one iteration of the cross-validation. You can visualize using the following code for ease. As a part of your submission, turn in an image of the generated tree.

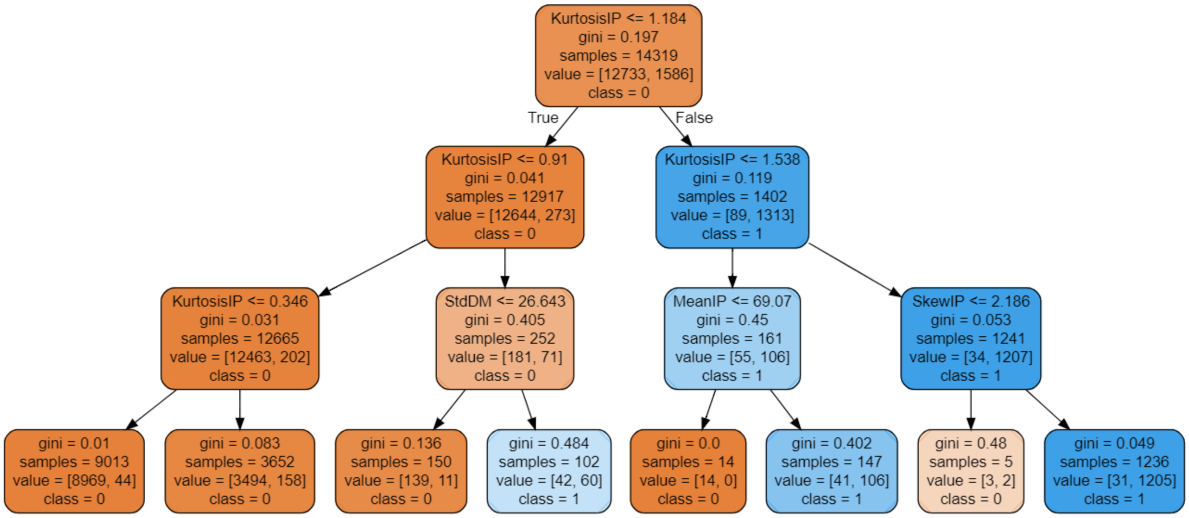
Step 1:

```
from sklearn import tree

Y_Classes = ['0','1']
tree.export_graphviz(<variable of the decision tree>, out_file =
'htru_tree.dot',
                     feature_names = <insert the list of feature names here>,
                     class_names = Y_Classes,
                     filled=True, rounded=True)
```

Step 2:

Copy paste the text generated in htru_tree.dot on http://www.webgraphviz.com/ to generate the decision tree.




Notice that majority of the class 0 samples fall in the left most leaf while majority of the class 1 samples in the right most tree. They are also the nodes with minimum gini index (low gini index implies better separation between classes).