National Taiwan University   **Introduction to Machine Learning and Deep Learning**

Department of Civil Engineering   Instructor: C.-S. CHEN

Homework 3

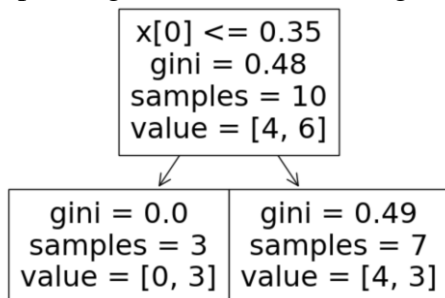**Essay and Programming, Due 21:00, Wednesday, October 23, 2024**

Total 120%

1. **(30%)** Consider three ensembles with 9, 25, 101 binary classifiers, respectively. Each ensemble has the same base classifier with an error rate of $\varepsilon_{\text{base}}$. The ensemble classifier predicts the class label of a test example by taking a majority vote on the prediction made by the base classifiers. Assume that the base classifiers are independent, and the ensemble makes a wrong prediction only if more than half of the base classifiers predict incorrectly. Write a Python program to compute $\varepsilon_{\text{ensemble}}$ and plot the relationship of $\varepsilon_{\text{base}}$ vs. $\varepsilon_{\text{ensemble}}$. Write your answer in `HW3_report_template.docx.`

2. **(30%)** A decision stump is a machine-learning model consisting of a one-level decision tree. Write a Python program named `HW3_2.ipynb` and classify the following dataset using a decision stump.

x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
y	1	1	1	-1	-1	-1	-1	1	1	1

Use `sklearn.tree.plot_tree` to plot the splitting result below and write a short essay explaining the contents of the figure. Write your answer in `HW3_report_template.docx.`

```
        x[0] <= 0.35
         gini = 0.48
       samples = 10
       value = [4, 6]
```
```
 gini = 0.0         gini = 0.49
samples = 3        samples = 7
value = [0, 3]     value = [4, 3]
```

3. **(60%)** Consider a dataset given in the following:

x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
y	1	1	1	-1	-1	-1	-1	-1	1	1

Like Bagging, we can pick different examples to obtain a new training set in AdaBoost. Supposed we have three rounds of boosting, and each round has the following training record respectively:

Boosting Round 1

x	0.1	0.4	0.5	0.6	0.6	0.7	0.7	0.7	0.8	0.9
y	1	-1	-1	-1	-1	-1	-1	-1	-1	1

Boosting Round 2

x	0.1	0.1	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3
y	1	1	1	1	1	1	1	1	1	1

Boosting Round 3

x	0.2	0.3	0.4	0.4	0.4	0.4	0.5	0.6	0.6	0.7
y	1	1	-1	-1	-1	-1	-1	-1	-1	-1

We can use a decision stump to classify the problem (See Problem2). Use AdaBoost and follow what we have covered in the toy example to compute ε_i, α_i and the updated weights. Finally, find the combined classifier H. Write your answer in HW3_report_template.docx.

Submission Format

Convert HW3_report_template.docx to HW3_report.pdf, then place HW3_report.pdf and HW3_2.ipynb into a folder named {yourStudentID}_HW3 and compress it into a ZIP file for upload to NTU COOL. Below are the file formats for upload.

```
R11521608_HW3.zip
├── HW3_2.ipynb
└── HW3_report.pdf
```