CSCI 416/516 Homework #3

DUE: November 15, 2024, at 11:59 pm

CSCI 416/516: Each Problem begins with an allocation of points, represented as [u pts/g pts]. If you are registered in CSCI 416, you can receive up to u pts on this Problem; if you are registered in CSCI 516, you can receive up to g pts on this Problem. The last Problem is optional for undergraduates (CSCI 416) but required for graduates (CSCI 516). **Write down which session you are in / are you a graduate or undergraduate student**.

Submission: You need to submit both your homework report (answers to the Problems) and your jupyter notebook code (for Problem 4) as PDF files through Blackboard. Please show the work on how you reach the conclusion for each question.

	Cloudy	Not Cloudy
Raining	50/100	2/100
Not Raining	20/100	28/100

- **Problem 1 [2pts/1pts]: Joint Entropy.**
  Suppose $X = \{\text{Raining, Not raining}\}$, $Y = \{\text{Cloudy, Not Cloudy}\}$, give the above Table. What is the joint entropy, $H(X, Y)$? Show your work on how the conclusion is reached.

- **Problem 2 [2pts/1pts]: Specific Conditional Entropy.**
  Following the setup from the previous question, what is the entropy of cloudiness Y, given that it is raining ($X = \text{Raining}$)? Show your work on how the conclusion is reached.

- **Problem 3 [2pts/2pts]: Information Gain.**
  How is the Information Gain defined? What is its relation to entropy?

- **Problem 4 [2pts/2pts]: AdaBoost.**
  Explain what this formula does, in the context of AdaBoost by answering: (1) what is this formula? (2) what does the part $y_i \log h_{\boldsymbol{\theta}}(\boldsymbol{x}_i) + (1 - y_i)\log(1 - h_{\boldsymbol{\theta}}(\boldsymbol{x}_i))$ do? (3) what does the part $\lambda ||\boldsymbol{\theta}_{[1:d]}||_2^2$ do? (4) What is the term $w_{i,t}$ and how does it affect $\mathcal{J}_{\text{reg},t}(\boldsymbol{\theta})$?

$$\mathcal{J}_{\text{reg},t}(\boldsymbol{\theta}) = -\sum_{i=1}^{n} w_{i,t}[y_i \log h_{\boldsymbol{\theta}}(\boldsymbol{x}_i) + (1 - y_i)\log(1 - h_{\boldsymbol{\theta}}(\boldsymbol{x}_i))] + \lambda ||\boldsymbol{\theta}_{[1:d]}||_2^2 \tag{1}$$

- **Problem 5 [2pts/2pts]: Decision Tree.**
  We want to train a decision tree model to classify breast cancer tumors into malignant

or benign categories using the Wisconsin Breast Cancer dataset. The Wisconsin Breast Cancer dataset contains measurements from digitized images of breast tissue samples. The measurements describe the characteristics of the cell nuclei present in the images. The objective is to classify tumors into benign (non-cancerous) or malignant (cancerous) based on these features.

- Load the Breast Cancer dataset using `sklearn.datasets.load_breast_cancer`.
- Split the dataset into training and testing sets using `train_test_split` with a test size of 30%.
- Scale the features using `StandardScaler`.
- Build a Decision Tree classifier using `DecisionTreeClassifier` from `sklearn.tree`.
- Fit the classifier on the training data.
- Predict the tumor classifications for the test data.
- Compute the accuracy of the model on the test data.
- Display the confusion matrix and classification report using `sklearn.metrics`.
- Visualize the decision tree using `sklearn.tree.plot_tree` or any other library of your choice like `graphviz`.

- **Problem 6 [bonus 2pts/2pts]: Conditional Entropy.**
We define the Conditional Entropy as

$$H(Y|X) = \sum_{x \in X} p(x) H(Y|X = x) \tag{2}$$

and then claim that

$$H(Y|X) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 p(y|x) \tag{3}$$

Show the process/work that leads to the claim/conclusion.