Midterm Exam
(Sample)

This sample midterm is provided for practice.
The real exam will include more questions and different topics.

1. **Short Questions**

   (a) **True** or **False**: Boosting can help reduce both bias and variance.

   (b) **True** or **False**: A more complex classifier is more likely to overfit than a simple classifier.

   (c) **True** or **False**: Cross validation can be used to select the number of iterations in boosting; this procedure may help reduce overfitting.

   (d) **True** or **False**: In SVMs, changing non-support vector data points will not change decision boundary, but we still need to use the entire training data to calculate the decision boundary.

   (e) What is the difference between a regression task and a classification task?

2. **Logistic Regression** We consider the following models for logistic regression for a binary classification ($y_i \in [0, 1]$):

$$f(\mathbf{x}) = \sigma(\mathbf{w}^\top \mathbf{x})$$

with loss function as:

$$L = \sum_i^N -\{y_i \log f(\mathbf{x}_i) + (1 - y_i) \log(1 - f(\mathbf{x}_i))\} + \lambda ||\mathbf{w}||^2$$

(a) Derive the gradient of L w.r.t $\mathbf{w}$

(b) Write down the update rule for $\mathbf{w}$

3. **Decision Trees** We will use the dataset below to learn a decision tree which predicts if people play outside (Yes or No), based on weather and temperature.

Weather	Temperature	Play Outside (Y)
Sunny	Hot	No
Sunny	Mild	Yes
Overcast	Hot	Yes
Rainy	Mild	Yes
Rainy	Cool	No
Sunny	Cool	Yes
Overcast	Mild	Yes
Rainy	Hot	No

For this problem, write your answer using $\log_2$, it may be helpful to note that $\log_2 3 \approx 1.6$.

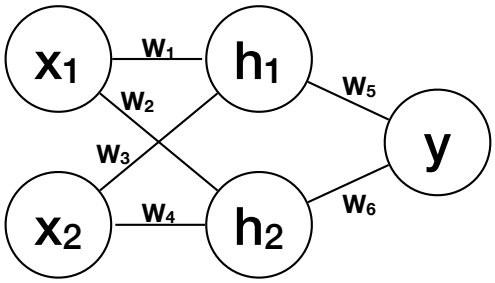
(a) What is the entropy $H(Y)$?

(b) What is the conditional entropy $H(Y|\text{Weather})$?

(c) What is the conditional entropy $H(Y|\text{Temperature})$?

(d) Describe what feature we should split on first.

4. **Neural Networks** Suppose we have a simplified Neural Network (shown below) with **<u>linear</u> activation units**.



(a) Can any function that is represented by the above network also be represented by a single unit neural network (or perceptron)? If so, draw this neural network, detailing the weights on the edges and the activation function. Otherwise, explain why not.