# CSE 417T (Machine Learning): Exam 2 Practice Questions

1. Assume we have a decision tree to classify binary vectors with 100 features. Can you specify a 1-Nearest Neighbor model that would result in exactly the same classification of all examples as the decision tree? Explain how, or provide a counter-example.

2. Consider the following dataset: the class variable is whether or not a car gets good mileage, and the features are Cylinders (either 4 or 8), Displacement (High, Medium, or Low), and Horsepower (High, Medium, or Low).

Cylinders	Displacement	Horsepower	GoodMPG?
4	Low	Low	No
8	High	High	No
8	High	Medium	Yes
8	High	High	No
4	Low	Medium	Yes
4	Medium	Medium	No

   (a) Give a decision tree that classifies this dataset perfectly.

   (b) What would be the first feature picked by ID3 to split on? Show your calculations.

3. Suppose you are using depth 1 decision trees (so you can split on one feature) as the weak learners in AdaBoost, but you know that all your features $x_i \in \{\pm 1\}$ (that is, you're solving a Boolean problem, and all features, as well as the class, are either true or false). Show that the final hypothesis output by AdaBoost after T rounds of boosting is a linear classifier.

4. Consider the following neural network with one hidden layer.
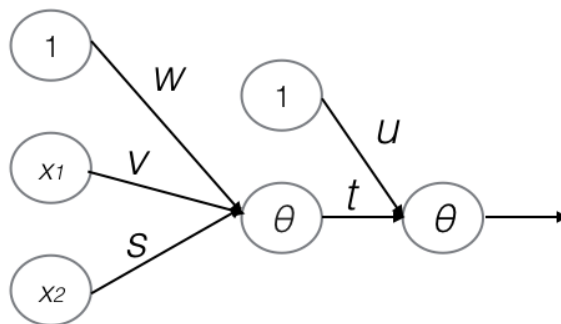
Assume we use the linear function $\theta(s) = s$ as the activation function for both the hidden layer and the output layer. Draw a neural network with no hidden layers (but with the same input and output layers) that is equivalent to the given one. You need to clearly specify the network structure and the weights.

5. Suppose you learn a hard-margin SVM using a training data set with N data points. You observe that there are M support vectors. Explain why the following bound on the leave-one-out cross validation error must be true: $E_{LOOCV} \leq M/N$

6. Suppose you apply bagging and boosting to linear classifiers. For which of the two will the output ensemble method also be a linear classifier?
   (a) Boosting only. **(b)** Bagging only. **(c)** Both boosting and bagging. **(d)** Neither boosting nor bagging.