

CS60020: Foundations of Algorithm Design and Machine Learning
Assignment 2
Submission deadline 29th Jan 2018

1. Consider the following set of training examples to train a machine to predict whether or not an office contains a recycling bin.

	STATUS	FLOOR	DEPT.	OFFICE SIZE	RECYCLING BIN?
1	faculty	four	cs	medium	Yes
2.	faculty	four	ee	medium	Yes
3.	student	four	cs	small	No
4.	faculty	five	cs	medium	Yes

Assume that each of the attributes assumes just the values that appear on the table.

- a) What is the size of the set of instances for this example?
- b) What is the size of the hypothesis space?
- c) Give a sequence of S and G boundary sets computed by the CANDIDATE-ELIMINATION algorithm if it is given the sequence of examples above in the order in which they appear on the table.
- d) Suggest a query guaranteed to reduce the size of the Version Space regardless of how the trainer classifies it.
2. a) You are given the following five training instances:
- $x_1 = 1, y_1 = 3.4$
 - $x_2 = 1.5, y_2 = 4.7$
 - $x_3 = 2, y_3 = 6.15$
 - $x_4 = 2.25, y_4 = 6.4$
 - $x_5 = 4, y_5 = 10.9$
- Using the derivation results for the parameters in ordinary least squares, calculate the values of b_0 and b .
- b) Suppose in a regression problem, our aim is to reduce the mean cubed error instead of mean squared error. Take the error function of the form $\sum_{i=1}^n (y_i - w^T x_i)^3$. Find the corresponding weights. Explain why we do not use mean cubed error and use mean squared error instead as an error function in regression.
3. What is the VC dimension of circles in a plane? Justify your answer in 4-5 sentences.

4. Choose the correct option(s) from the following explaining the reason for choosing or rejecting each option in 1-2 sentences for each option:
- a) When working with a small dataset, one should prefer low bias/high variance classifiers over high bias/low variance classifiers.
 - b) When working with a small dataset, one should prefer high bias/low variance classifiers over low bias/high variance classifiers.
 - c) When working with a large dataset, one should prefer high bias/low variance classifiers over low bias/high variance classifiers.
 - d) When working with a large dataset, one should prefer low bias/high variance classifiers over high bias/low variance classifiers.