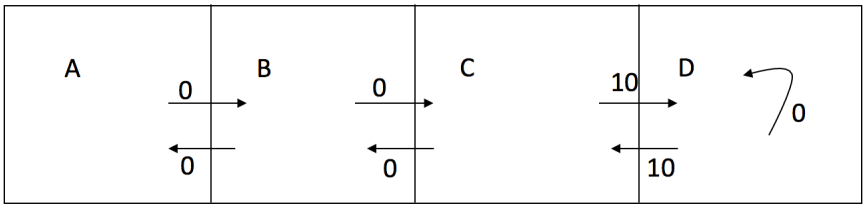
COMP 7745/8745: Machine Learning

Instructor: Deepak Venugopal

Spring 2018: Homework 3

Due Date: April 8, 2018 (Hard copy in class or softcopy in ecourseware)

1. For the below grid world, what is the i) optimal policy ii) Approximate Q-values that the Q-learning algorithm will converge to assuming a discount factor of 0.1 (you should be able to do this without much calculation or running the Q-learning algorithm). (10 points)



2. If the discount factor is equal to 0, then is it harder or easier to learn the optimal policy. Briefly explain. (5 points)

3. Your boss says, "you should only use a classifier it it gives you 100% accuracy on the training dataset". How will you counter his/her argument using the bias vs variance tradeoff? Briefly explain. (10 points)

4. In the ADABoosting algorithm, suppose in iteration t+1, we increase the weights of those data points that are correctly classified in iteration t. Is this a good idea. Explain your reasoning. (10 points)

5. For each of the following, state whether they have high/low bias and high/low variance. a) Perceptron on linearly separable data b) Neural networks on not-linearly separable data c) K-NN on linearly separable data d) Perceptrons on not-linearly separable data (10 points)

6. An SVM is trained with the following data (10 points)

X_0	X_1	y
-1	-1	-1
1	1	1
0	2	1

a) Show the kernel matrices for i) the linear kernel ii) polynomial kernel of degree 3 b) Assuming that the Lagrangian coefficients for the SVM optimization problem have the values, $\alpha_1 = 1/8$, $\alpha_2 = 1/8$, $\alpha_3 = 0$ Which data instances are the support vectors?

7. Given the following dataset, we wish to apply Adaboosting using SVMs with linear kernels as our base classifier. Is this a good idea? Explain your answer. (10 points)

X_0	X_1	y
0	0	-1
0	1	-1
1	0	-1
1	1	1

8. Given the following dataset, which kernel for the support vector machine would you pick and why?(10 points)

X_0	X_1	y
0	0	-1
0	1	1
1	0	1
1	1	1

9. Given a support vector machine trained on m examples, what is an upper bound on its leave-one-out cross validation error given that it has a) m support vectors b) $m/2$ support vectors (5 points)

10. Here, you will experiment with the SVM implementation in Weka for the given dataset and Adaboosting with the wines dataset. For this, download the LIbSVM jar (provided in the zip file). You need to start Weka GUI from commandline with the libsvm.jar in the class path. E.g. java -classpath weka.jar;libsvm.jar weka.gui.GUIChooser. The SVM implementation shows up as SMO under functions, and Adaboosting is under Meta. (20 points)

    Report the 10-fold cross-validation results (average precision,recall and F1) for each of the 3 kernel types (linear (polynomial degree 1), RBF and polynomial degree 2) for the cost factors, 1, 10, 100 and 1000. Briefly explain your observations of the effect of kernel types and cost factors on computation time and accuracy?

    For Adaboosting, you will experiment with using weak and strong base classifiers. For the first-case, choose decision stumps as the base classifier, and for the second case use J48 as the base classifier. Does Adaboosting help boost the performance of both? Specifically, compare the 10-fold cross-validation results (average precision,recall and F1) for the base classifiers (decision stumps and J48), with the Adaboosted versions of these base classifiers.