CSE 4088 – INTRODUCTION to MACHINE LEARNING

HW2 Report

Zeynep Kumralbaş 150115051

1) $2Me^{-2\varepsilon^2 N} \leq 0.03 \qquad e^{-2\varepsilon^2 N} \leq \frac{0.03}{2M} \qquad -2\varepsilon^2 N \leq \ln\left(\frac{0.03}{2M}\right) \qquad N \leq \frac{\ln\left(\frac{0.03}{2M}\right)}{-2\varepsilon^2}$

$\varepsilon = 0.05, M = 1 \qquad N \geq 839.94$

The least number of examples N among the given choices is 1000. The answer is b.

2) $\varepsilon = 0.05, M = 10 \qquad N \geq 1300$

The least number of examples N among the given choices is 1500. The answer is c.

3) $\varepsilon = 0.05, M = 100 \qquad N \geq 1760.97$

The least number of examples N among the given choices is 2000. The answer is d.

The Perceptron Learning Algorithm

Step1: Firstly, two points are selected randomly. The target function *f* is constructed taking the line passing through these two points.

Step2: *N* random points are selected randomly, and their signs are decided based on target function *f*. The sign of a point is determined +1 if it is in the upper side of the *f*, otherwise -1.

Step3: Initial *w* is 0. In the first iteration of the perceptron learning algorithm all points are misclassified.

Step4: One misclassified point is chosen randomly, and *w* is updated.

Step5: Then, *g* is evaluated with the new *w*, misclassified points are detected.

Step 4 and step 5 continue to be executed until the convergence occurs.

4) This experiment is run 1000 times with *N*=10. Average number of iterations for the PLA to converge is 11.0490. This result is close to 15. The answer is b.

5) 1000 random fresh points are created randomly. The probability that *f* and *g* will disagree on their classification is evaluated using these fresh points. The average number of misclassified points are 0.1037. This result is close to 0.1. The answer is c.

6) The experiment is run 1000 times with *N*=100. Average number of iterations for the PLA to converge is 98.7870. This result is close to 100. The answer is b.

7) 1000 random fresh points are created randomly. The probability that *f* and *g* will disagree on their classification is evaluated using these fresh points. The average number of misclassified points are 0.0139. This result is close to 0.01. The answer is b.

Linear Regression

Step1: Firstly, two points are selected randomly. The target function *f* is constructed taking the line passing through these two points.

Step2: *N* random points are selected randomly, and their signs are decided based on target function *f*. The sign of a point is determined +1 if it is in the upper side of the *f*, otherwise -1.

Step3: $X^{\dagger} = (X^T X)^{-1} X^T$ and $w = X^{\dagger} y$ is calculated.

8) The experiment is run 1000 times with *N*=100. The average (averageEin) $E_{in} = 0.0398$. This result is close to 0.01. The answer is c.

9) 1000 random fresh points are created randomly. The average (averageTotalFreshMisclassified) $E_{out} = 0.0485$. This result is close to 0.01. The answer is c.

10) The *w* is found using Linear Regression algorithm with *N*=10. Then, this *w* is given as initial weights for the Perceptron Learning Algorithm. The average number of iterations that PLA takes to converge is 5.4620. This result is close to 1. The answer is a.

Nonlinear Transformation

Step1: N points are created randomly, and their signs are determined based on the given target function f.

Step2: Signs of the 10% of the data points are flipped to generate noise.

11) Linear Regression algorithm runs, *w* is found. True signs of data points and evaluated signs using *w*, are compared. The experiment is run 1000 times. Average in-sample error (averageEin) $E_{in} = 0.5051$. This result is close to 0.5. The answer is d.

12) Nonlinear feature vector is created for N=1000 data point. Linear Regression algorithm is used to find *w*. The experiment is run 10 times to obtain the average *w*. The result of *w* is:

(-1,02635431313071, -0,0388484313084285, -0,0320520387097557, -0,0450079235123249, 1,62514080852391, 1,61511443209175)

The hypothesis that is the closest to found one is in (a). The answer is a.

13) 1000 random fresh points are created randomly, signs of the 10% data points are flipped to generate noise. The experiment is run 1000 times. The average (averageTotalFreshMisclassified) $E_{out} = 0.1257$. This result is close to 0.1. The answer is b.