كلية الحاسبات والذكاء الإصطناعي
Faculty of Computers & Artificial Intelligence

جامعة حلوان

Name:...

ID:...............

CS395: Selected Topics in CS-1	Number of pages: 3
Course Coordinator: Assoc.Prof. Ghada Khoriba ,	Total Marks: 10
Date: 30/11/2020 Time:3:00-3:45	

[Model A B C D] Choose the best answer:

1.Suppose you are working on weather prediction, and use a learning algorithm to predict tomorrow's temperature (in degrees Centigrade/Fahrenheit).Would you treat this as a classification or a regression problem?

a.Regression

b.classification

2.Consider the following training set of m= 4 training examples: x ={ 1, 2 , 4, 0} y= {0.5,1 ,2,0}. Consider the linear regression model $h_\theta(x)=\theta_0+\theta_1x$. What are the values of θ_0 and θ_1 that you would expect to obtain upon running gradient descent on this model?

a.0.5,0.5

b.0.5,0

c.1,1

d.0,0.5

3.In the previous problem, Suppose we set $\theta_0=-1,\theta_1=0.5$. What is $h_\theta(4)$?

a.0

b.1

c.3

d.2

4.Suppose you have a dataset with m=500 examples and n=20 features for each example. You want to use multivariate linear regression to fit the parameters θ to our data. Should you prefer gradient descent or the normal equation?

a.Gradient descent, since $(X^TX)^{-1}$ will be very slow to compute in the normal equation.

b.The normal equation, since it provides an efficient way to directly find the solution.

c.Gradient descent, since it will always converge to the optimal θ.

d.The normal equation, since gradient descent might be unable to find the optimal θ.

5.Suppose we use gradient descent to try to minimize $f(\theta_0,\theta_1)$ as a function of θ_0 and θ_1.If θ_0 and θ_1 are initialized at a local minimum, then one iteration will change their values by learning rate α.

a.True

b.False

6.Suppose we use gradient descent to try to minimize $f(\theta_0,\theta_1)$ as a function of θ_0 and θ_1.Even if the learning rate α is very large, every iteration of gradient descent will decrease the value of $f(\theta_0,\theta_1)$.

a.True

b.False

7.Suppose you have m=20 training examples with n=6 features (excluding the additional all-ones feature for the intercept term, which you should add). The normal equation is $\theta=(X^TX)^{-1}X^Ty$. For the given values of m and n, what are the dimensions of θ, X, and y in this equation?

a.X is 20×6, y is 20×1, θ is 6x6

b.X is 20×7, y is 20×1, θ is 7×1

c.X is 20×6, y is 20×1, θ is 7×1

d.X is 20×7, y is 20×7, θ is 7×7

8.When using feature scaling, It speeds up gradient descent by making it require fewer iterations to get to a good solution.

a.True

b.False

9.In logistic regression, adding a new feature to the model

a.always result in equal performance

b.always result in equal or better performance on the training set but may lead to overfitting

c.adding many new features to the model helps prevent overfitting on the training set.

d.always result in bad performance

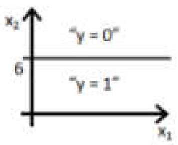
Name:..

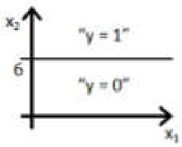
ID:...............

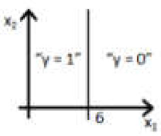
10. The hypothesis follows the data points very closely and is highly complicated, indicating that it is overfitting the training set.
    - a. True
    - b. False

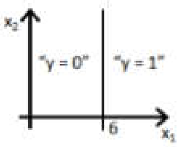
11. Suppose you ran logistic regression, with $\lambda = 1$, and once with $\lambda = 0$. which value of θ corresponds to $\lambda = 0$.
    - a. $\theta = 11.02$, 0.25
    - b. $\theta = 91.02$, 12.25

12. Regularized logistic regression and regularized linear regression are both convex, and gradient descent will still converge to the global minimum.
    - a. True
    - b. False

13. suppose we train a logistic regression $h_\theta = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$, suppose $\theta_0 = 6$, $\theta_1 = 0$, $\theta_2 = -1$, which of the following represents the decision boundary found by the classifier.
    - a.


    - b.


    - c.


    - d.


14. Using a very large value λ (regularization term) can hurt the performance of your hypothesis; the only reason we do not set to be too large is to avoid numerical problems.
    - a. True
    - b. False

15. If we trained a logistic regression classifier, and it outputs on a new example x a prediction $h_\theta(x) = 0.8$ this means
    - a. our estimation for $P(y=0|x;\theta)$ is 0.8
    - b. our estimation for $P(y=0|x;\theta)$ is 0.2
    - c. our estimation for $P(y=1|x;\theta)$ is 0.8
    - d. our estimation for $P(y=1|x;\theta)$ is 0

16. If the learning rate is too small, then gradient descent may take a very long time to converge.
    - a. True
    - b. False

17. Suppose m=4 students have taken some class, and the class had a midterm exam and a final exam. You have collected a dataset of their scores on the two exams, which is as follows:
    You'd like to use polynomial regression to predict a student's final exam score from their midterm exam score. Concretely, suppose you want to fit a model of the form $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$, where x_1 is the midterm score and x_2 is (midterm score)^2. Further, you plan to use both feature scaling and mean normalization. What is the normalized feature $x_2^{(4)}$?

Midterm Exam	(midterm exam)^2	Final Exam
89	7921	96
72	5184	74
94	8836	87
69	4761	78

    - a. 0.47
    - b. 0.36
    - c. 6.6
    - d. -0.47

Name:...
ID:...............

18. Given a large dataset of medical records from patients suffering from heart disease, try to learn whether there might be different clusters of such patients for which we might tailor separate treatments.
   a. best addressed using a supervised learning algorithm
   ● b. best addressed using an unsupervised learning algorithm.

19. By the definition of $J(\theta_0,\theta_1)$, it is not possible for there to exist θ_0 and θ_1 so that $J(\theta_0,\theta_1)=0$
   a. True
   ● b. False

20. If $J(\theta_0,\theta_1)=0$ that means the line defined by the equation "$y = \theta_0 + \theta_1 x$" perfectly fits all of our data. the values of θ_0 and θ_1 that achieve this are both 0.
   a. it is not possible for there to exist θ_0 and θ_1 so that $J(\theta_0,\theta_1)=0$
   ● b. $y(i)=0$ for all of training examples
   c. $y(i)=1$ for all of training examples
   d. $h_\theta(x)= x$