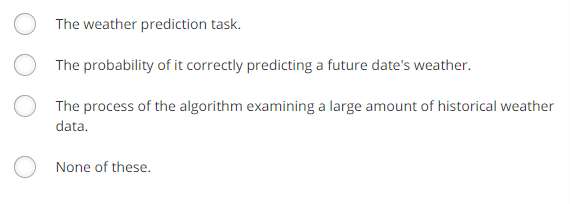
# 引论

1 A computer program is said to learn from experience E with respect to some task T and some performance measure P if its performance on T, as measured by P, improves with experience E. Suppose we feed a learning algorithm a lot of historical weather data, and have it learn to predict weather. What would be a reasonable choice for P?

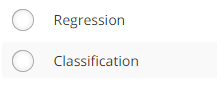


2 Suppose you are working on weather prediction, and you would

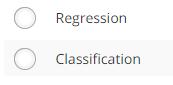
like to predict whether or not it will be raining at 5pm

tomorrow. You want to use a learning algorithm for this.

Would you treat this as a classification or a regression problem?



3 Suppose you are working on stock market prediction. You would like to predict whether or not a certain company will declare bankruptcy within the next 7 days (by training on data of similar companies that had previously been at risk of bankruptcy). Would you treat this as a classification or a regression problem?



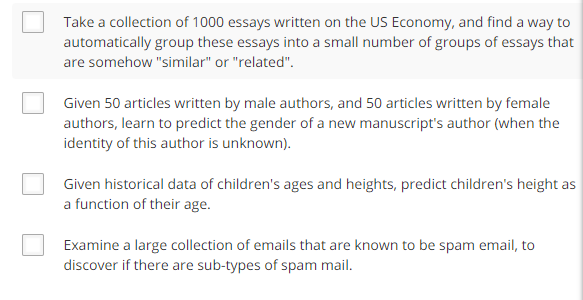
4 Some of the problems below are best addressed using a supervised

learning algorithm, and the others with an unsupervised

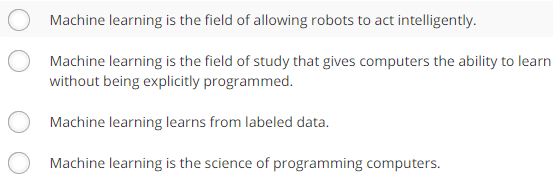
learning algorithm. Which of the following would you apply

supervised learning to? (Select all that apply.) In each case, assume some appropriate

dataset is available for your algorithm to learn from.



5 Which of these is a reasonable definition of machine learning?

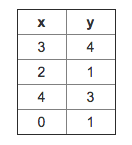


# Linear Regression with One Variable

1 Consider the problem of predicting how well a student does in her second year of college/university, given how well she did in her first year.

Specifically, let x be equal to the number of "A" grades (including A-. A and A+ grades) that a student receives in their first year of college (freshmen year). We would like to predict the value of y, which we define as the number of "A" grades they get in their second year (sophomore year).

Refer to the following training set of a small sample of different students' performances (note that this training set may also be referenced in other questions in this quiz). Here each row is one training example. Recall that in linear regression, our hypothesis is hθ(x)=θ0+θ1x, and we use m to denote the number of training examples.



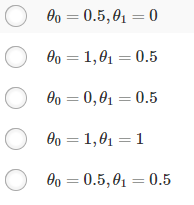
For the training set given above, what is the value of *m*? In the box below, please enter your answer (which should be a number between 0 and 10).

( 在此处填空 )

2 Consider the following training set of *m*=4 training examples:

|  |  |
| --- | --- |
| x | y |
| 1 | 0.5 |
| 2 | 1 |
| 4 | 2 |
| 0 | 0 |

Consider the linear regression model *hθ*(*x*)=*θ*0+*θ*1*x*. What are the values of *θ*0 and *θ*1 that you would expect to obtain upon running gradient descent on this model? (Linear regression will be able to fit this data perfectly.)



3 Suppose we set *θ*0=−1,*θ*1=2 in the linear regression hypothesis from Q1. What is *hθ*(6)?

（在此处填空）

4 Let *f* be some function so that

*f*(*θ*0,*θ*1) outputs a number. For this problem,

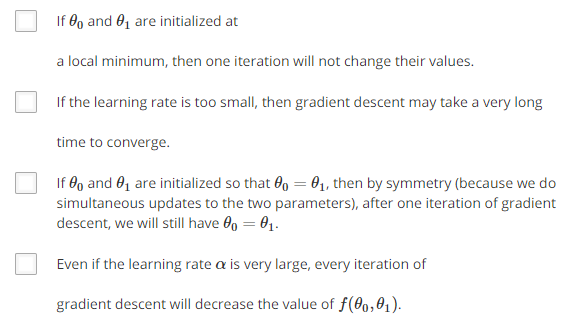
*f* is some arbitrary/unknown smooth function (not necessarily the

cost function of linear regression, so *f* may have local optima).

Suppose we use gradient descent to try to minimize *f*(*θ*0,*θ*1)

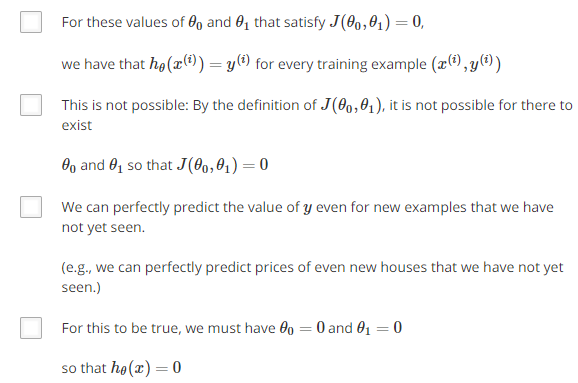
as a function of *θ*0 and *θ*1. Which of the

following statements are true? (Check all that apply.)



5 Suppose that for some linear regression problem (say, predicting housing prices as in the lecture), we have some training set, and for our training set we managed to find some *θ*0, *θ*1 such that *J*(*θ*0,*θ*1)=0.

Which of the statements below must then be true? (Check all that apply.)



# 线性代数回顾（省略）·