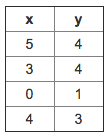
Linear Regression with One Variable

5 questions

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

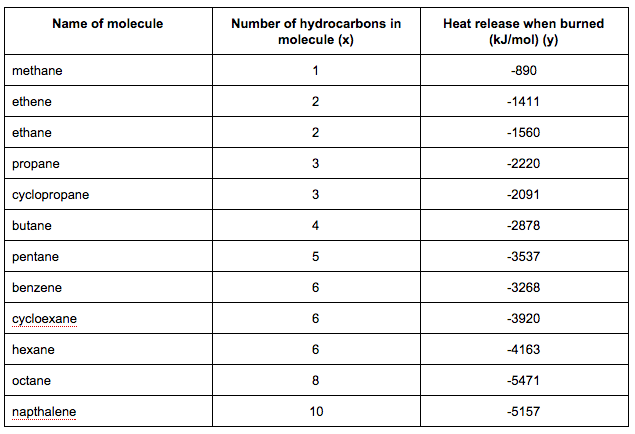
Questions 1 through 4 will use the following training set of a small sample of different students' performances. Here each row is one training example. Recall that in linear regression, our hypothesis is*hθ*(*x*)=*θ*0+*θ*1*x*, and we use *m* to denote the number of training examples.



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).



2. Many substances that can burn (such as gasoline and alcohol) have a chemical structure based on carbon atoms; for this reason they are called hydrocarbons. A chemist wants to understand how the number of carbon atoms in a molecule affects how much energy is released when that molecule combusts (meaning that it is burned). The chemists obtains the dataset below. In the column on the right, “kJ/mol” is the unit measuring the amount of energy released.



You would like to use linear regression (*hθ*(*x*)=*θ*0+*θ*1*x*) to estimate the amount of energy released (y) as a function of the number of carbon atoms (x). Which of the following do you think will be the values you obtain for *θ*0 and *θ*1? You should be able to select the right answer without actually implementing linear regression.



*θ*0=−569.6,*θ*1=530.9



*θ*0=−1780.0,*θ*1=530.9



*θ*0=−1780.0,*θ*1=−530.9



*θ*0=−569.6,*θ*1=−530.9

3. Suppose we set *θ*0=0,*θ*1=1.5. What is *hθ*(2)?



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.)

If *θ*0 and *θ*1 are initialized at

a local minimum, then one iteration will not change their values.

Even if the learning rate *α* is very large, every iteration of

gradient descent will decrease the value of *f*(*θ*0,*θ*1).

If the learning rate is too small, then gradient descent may take a very long

time to converge.

If *θ*0 and *θ*1 are initialized so that *θ*0=*θ*1, then by symmetry (because we do simultaneous updates to the two parameters), after one iteration of gradient descent, we will still have *θ*0=*θ*1.

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.)

For this to be true, we must have *θ*0=0 and *θ*1=0

so that *hθ*(*x*)=0

This is not possible: By the definition of *J*(*θ*0,*θ*1), it is not possible for there to exist

*θ*0 and *θ*1 so that *J*(*θ*0,*θ*1)=0

For these values of *θ*0 and *θ*1 that satisfy *J*(*θ*0,*θ*1)=0,

we have that *hθ*(*x*(*i*))=*y*(*i*) for every training example (*x*(*i*),*y*(*i*))

We can perfectly predict the value of *y* even for new examples that we have not yet seen.

(e.g., we can perfectly predict prices of even new houses that we have not yet seen.)