4 0.5 3

NO No matter how θ_0 and θ_1 are initialized, so long as α is sufficiently small, we can safely expect gradient descent to converge to the same solution. Correct 0.25

This is not true, because depending on the initial condition, gradient descent may end u

YES If the learning rate is too small, then gradient descent may take a very long time t
Correct 0.25
If the learning rate is small, gradient descent ends up taking an extremely small st

YES If θ_0 and θ_1 are initialized at the global minimum, the one iteratio
Correct 0.25
At the global minimum, the derivative (gradient) is zero, so gradient descent will n

NO Setting the learning rate α to be very small is not harmful, and can only spee
Correct 0.25
If the learning rate is small, gradient descent ends up taking an extremely small ste

If ?0 and ?1 are initialized at a local minimum, the one iteration will not
change their values. Inorrect 0.00
At a local minimum, the derivative (gradient) is zero, so gradient descent
will not change the parameters.

YES If the first few iterations of gradient descent cause $f(\theta_0, \theta_1)$ to increase rather than decrease, then the most likely cause is that we have set the learning rate a to too large a value. Inorrect 0.00 If alpha were small enough, then gradient descent should always successfully take a tiny small downhill and decrease $f(\theta_0, \theta_1)$ at least a little bit.

If gradient descent instead increases the objective value, that means alpha is too large

YES If ?0 and ?1 are initialized at the global minimum, the one iteration will not chang
Correct 0.25 At the global minimum, the derivative (gradient) is zero, so gradient desce
YES No matter how ?0 and ?1 are initialized, so long as a is sufficiently small, we can
Correct 0.25 This is not true, because depending on the initial condition, gradient desc

NO Even if the learning rate a is very large, every iteration of gradient descent will d
Inorrect 0.00 If the learning rate a is too large, one step of gradient descent can actu

YES If ?0 and ?1 are initialized so that ?0=?1, then by symmetry (because we do simultan
Inorrect 0.00 The updates to ?0 and ?1 are different (even though we're doing simultaneo

Suppose that for some linear regression problem (say, predicting housing prices as in th
managed to find some θ_0, θ_1 such that $J(\theta_0, \theta_1)$=0. Which o

Your Answer Score Explanation

```
NO We can perfectly predict the value of y even for new examples that we have not yet se
Inorrect 0.00 Even though we can fit our training set perfectly, this does not mean that

NO This is not possible: By the definition of J(?0,?1), it is not possible for there to 
Correct 0.25 If all of our training examples lie perfectly on a line, then J(?0,?1)=0 is

YES Our training set can be fit perfectly by a straight line, i.e., all of our training 
Inorrect 0.00 If J(?0,?1)=0, that means the line defined by the equation "y=?0+?1x" perf

NO Gradient descent is likely to get stuck at a local minimum and fail to find the globa
Inorrect 0.00 The cost function J(?0,?1) for linear regression has no local optima (othe

NO For this to be true, we must have y(i)=0 for every value of i=1,2,,m.
Correct 0.25 So long as all of our training examples lie on a straight line, we will be 
NO For this to be true, we must have $\theta_0$=0 and $\theta_1$=0 so that $h_\theta$(x)
Correct 0.25
```

If $J(\theta_0, \theta_1)$=0, that means the line defined by the equation "y=θ_0+θ_1x" perfectly fits all of our data. There's no particular reason to expect that the values of θ_0 and θ_1 that achieve this are both 0 (unless y(i)=0 for all of our training examples).

NO This is not possible: 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 Correct 0.25 If all of our training examples lie perfectly on a line

```
, then $J(\theta_0, \theta_1)$=0 is possible.

YES Gradient descent is likely to get stuck at a local minimum and fail to find the glob
   Incorrect 0.00 The cost function $J(\theta_0, \theta_1)$ for linear regression has no
   so gradient descent will not get stuck at a bad local minimum.

NO Our training set can be fit perfectly by a straight line, i.e., all of our training e
   Incorrect 0.00
   If $J(\theta_0, \theta_1)$=0, that means the line defined by the equation "y=$\theta_
   of our data.

YES For these values of ?0 and ?1 that satisfy J(?0,?1)=0, we have that h?(x(i))=y(i) fo
Inorrect 0.00 J(?0,?1)=0, that means the line defined by the equation "y=?0+?1x" perfect
```