1. Draw the space of the following equation:

f(x1, x2) = 3x1 - 4x2 + 5x12 + 2x1x2 + 7x22 , when the range of the variables

are: -12 ≤ x1, x2 ≤ 12. Provide the drawing as well as the code of your drawing.

b) Is it a convex function or a concave function? Explain, ‘why’ in terms of the given

equation.

(c) Does local minimum exist – Yes/No?

1. Which linear regression training algorithm can you use If you have a training set with millions of features?
2. Suppose the features in your training set have very different scales.
3. Which algorithms might suffer from this, and how?
4. What can you do about it?
5. Do all Gradient Descent algorithms lead to the same model, provided you let them run long enough?
6. Suppose you use Batch Gradient Descent and you plot the validation(or,test) error at every epoch. (a) If you notice that the validation error consistently goes up, what is likely going on? What if the training error is (b) either going up or (c) going down at the same time? (d) How can you fix this?
7. (a) Which Gradient Descent algorithm (among those we discussed) will reach the vicinity of the optimal solution the fastest? (b) Which will actually converge? (c) How can you make the others converge as well? Explain.