

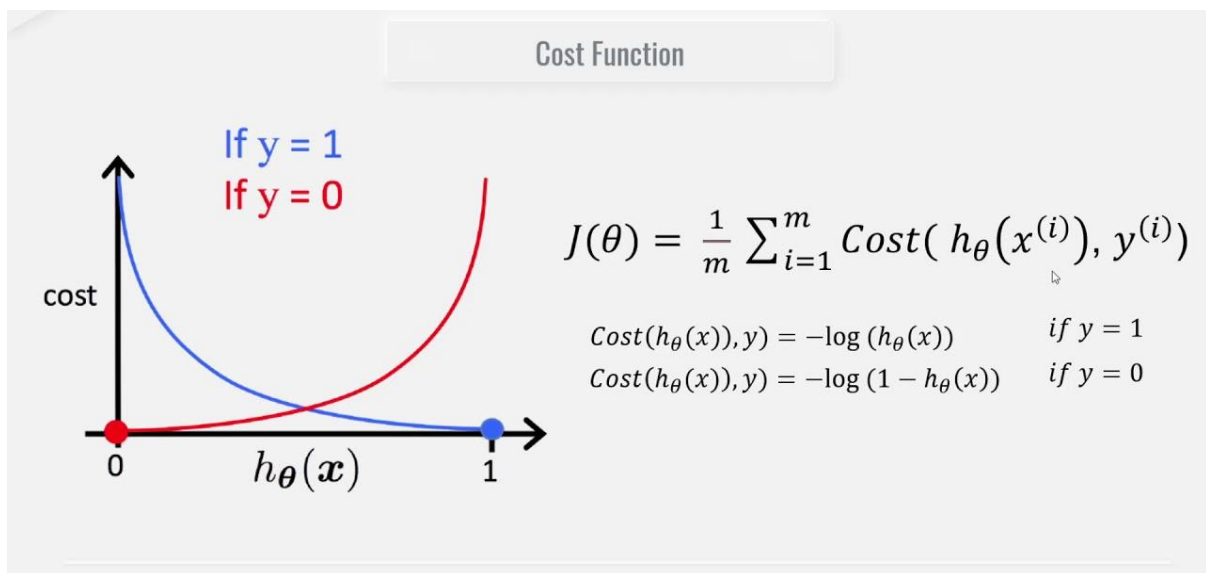
Assignment 5

1. Which Linear Regression training algorithm can you use if you have a training set with millions of features?

Linear Regression having training sets with millions of features can use **batch gradient descent, stochastic gradient descent, or mini-batch gradient descent.**

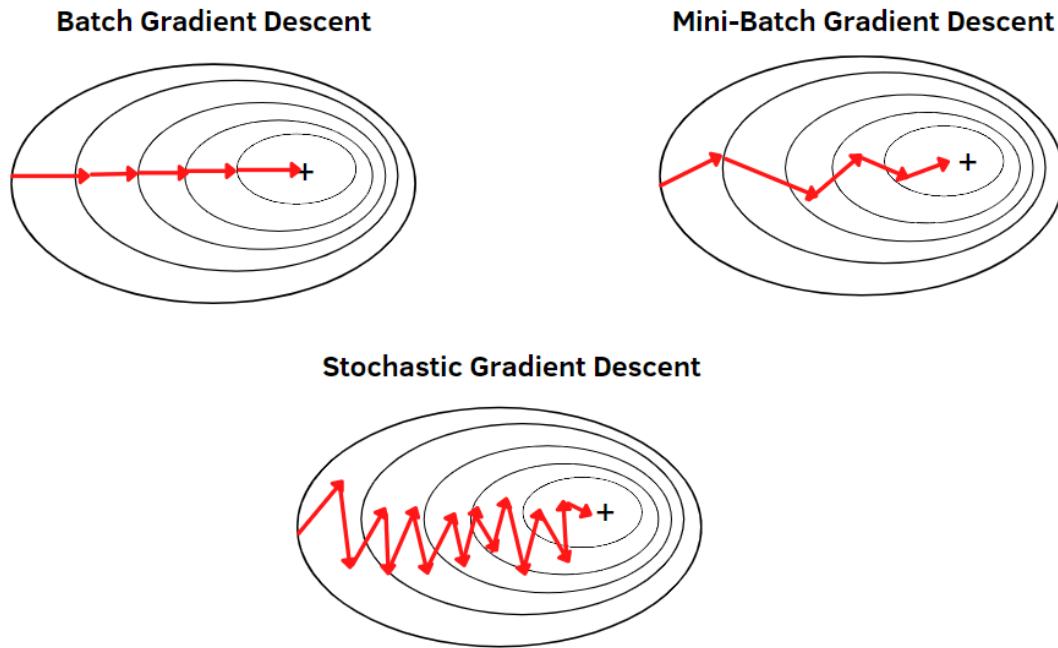
2. Can Gradient descent get stuck in a local minimum when training a logistic regression model?

Gradient Descent cannot get stuck in a local minimum when training a Logistic Regression model because the cost function is convex.



3. Do all Gradient Descent algorithms lead to the same model, provided they run for the same no. of epochs?

No. The issue is that **stochastic gradient descent and mini-batch gradient descent** have randomness built into them. This means that they can find their way to nearby the global optimum, but they generally don't converge



Therefore, they do not give same results even if they are run for the same number of epochs.

4. Suppose you are using a Polynomial Regression and you plot the learning curves and you notice there is a large gap between training and validation error. What is the problem? How to solve it?

When you plot learning curves and notice there is a large gap between the training error and the validation error that is characteristic of an over fitting model. The "gap" exists simply because validation error is more than training error.

- One way to improve an over fitting model is to provide more training data.
- Another way is to reduce the complexity of the model.
- You can also reduce the number of features in your data.
- One last thing to try is add regularization to your model. Either L2 (ridge regression) or L1 (lasso) are good choices.

5. Suppose the features in your training dataset are in different scales. Which algorithms will suffer from this? How to handle this situation?

Feature scaling is required for the various gradient descent algorithms. Feature scaling will help gradient descent converge quicker.

Feature scaling can be done using Minmax Scaler and Standard Scaler function to scale down the features and reduce complexity.