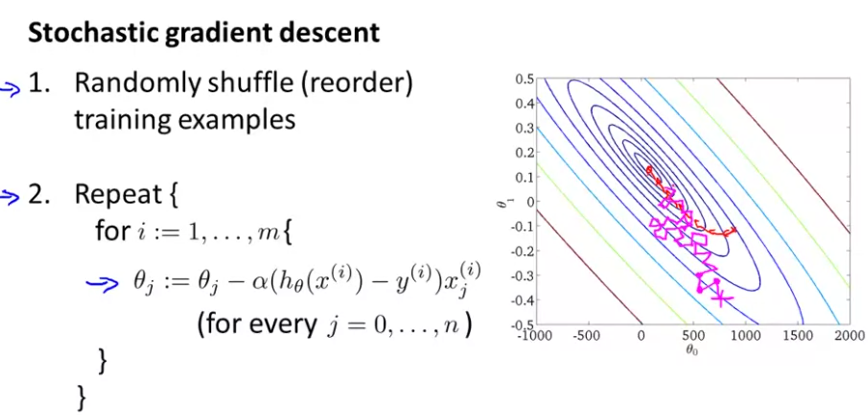
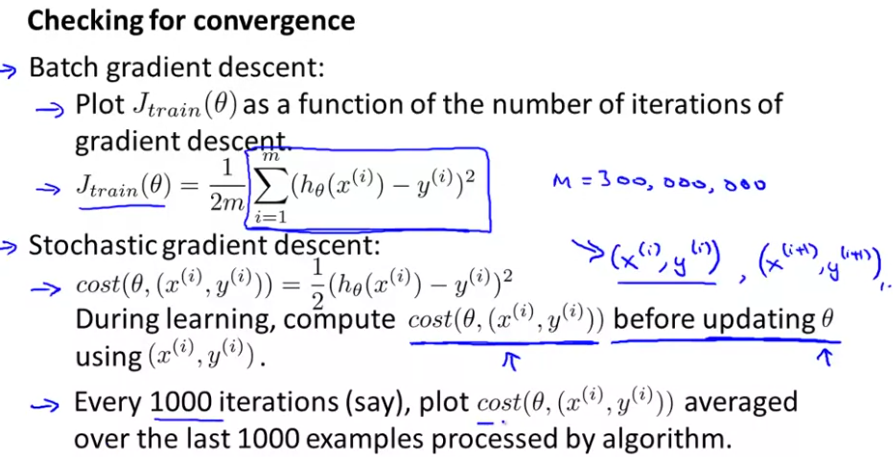
**Learning with large Datasets**

It is important to collect a great amount of data, as a low bias algorithm will actually work fine if it has enough data.

**Stochastic Gradient Descent**

The difference with respect to the classic gradient descent consists that instead of iterating many times over all the examples, we iterate only once over all of the examples.



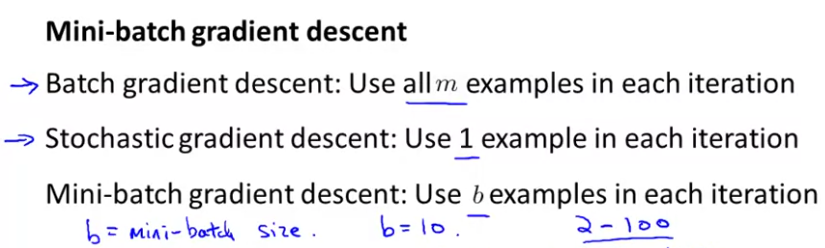


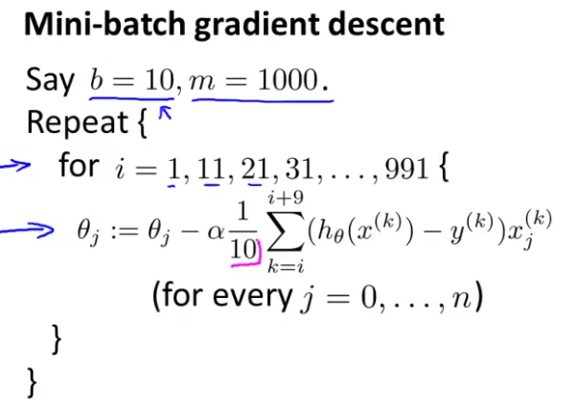
The stochastic gradient descent can converge to a local optima, and not to the global optima. Therefore, decreasing the parameter alfa can improve the algortihm’s performance.

When plotting the cost over the iterations, we can obtain a smoother plot if we average over a greater number of examples.

If you want stochastic gradient descent to actually converge to the global minimum, you can slowly decrease the learning rate alpha over time.

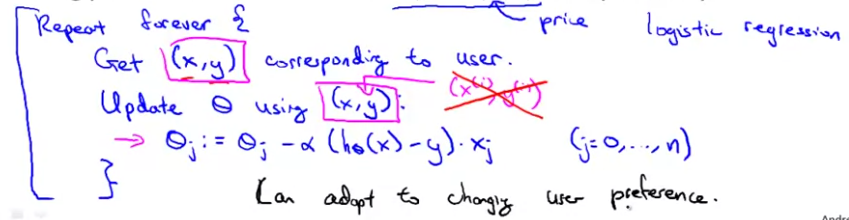
**Mini-batch gradient descent**





**Online Learning**

It allows us to model problems where we have a continuous flood or a continuous stream of data coming in and we would like an algorithm to learn from that. Specifically, if you have a continuous stream of data generated by a continuous stream of users coming to your website, what you can do is use an online learning algorithm to learn user preferences from the stream of data and use that to optimize some of the decisions on your website.



**Map reduce and data parallelism**

It’s useful for problems where our learning algorithms can be expressed as sums over the training set and that are so big that one only machine can’t handle the amount of data. It consists in splitting the training set in many pieces to process it in different machines at the same time:

