Large Scale Machine Learning

TOTAL POINTS 5

---

1. Suppose you are training a logistic regression classifier using stochastic gradient descent. You find that the cost (say, $cost(\theta,(x^{(i)},y^{(i)}))$, averaged over the last 500 examples), plotted as a function of the number of iterations, is slowly increasing over time. Which of the following changes are likely to help?

1 point

- ( ) Try averaging the cost over a larger number of examples (say 1000 examples instead of 500) in the plot.
- ( ) This is not an issue, as we expect this to occur with stochastic gradient descent.
- ( ) Try using a larger learning rate α.
- (•) Try using a smaller learning rate α.

2. Which of the following statements about stochastic gradient

   descent are true? Check all that apply.

1 point

- [x] In each iteration of stochastic gradient descent, the algorithm needs to examine/use only one training example.
- [x] One of the advantages of stochastic gradient descent is that it can start progress in improving the parameters θ after looking at just a single training example; in contrast, batch gradient descent needs to take a pass over the entire training set before it starts to make progress in improving the parameters' values.
- [ ] Suppose you are using stochastic gradient descent to train a linear regression classifier. The cost function $J(\theta)=\frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)})-y^{(i)})^2$ is guaranteed to decrease after every iteration of the stochastic gradient descent algorithm.
- [ ] Stochastic gradient descent is particularly well suited to problems with small training set sizes; in these problems, stochastic gradient descent is often preferred to batch gradient descent.

3. Which of the following statements about online learning are true? Check all that apply.

1 point

- [ ] One of the advantages of online learning is that there is no need to pick a learning rate α.

- [ ] One of the disadvantages of online learning is that it requires a large amount of computer memory/disk space to store all the training examples we have seen.

- [ ] In the approach to online learning discussed in the lecture video, we repeatedly get a single training example, take one step of stochastic gradient descent using that example, and then move on to the next example.

- [x] When using online learning, in each step we get a new example (x, y), perform one step of (essentially stochastic gradient descent) learning on that example, and then discard that example and move on to the next.

4. Assuming that you have a very large training set, which of the

following algorithms do you think can be parallelized using

map-reduce and splitting the training set across different

machines? Check all that apply.

1 point

- [x] Computing the average of all the features in your training set $\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}$ (say in order to perform mean normalization).

- [x] Logistic regression trained using batch gradient descent.

- [ ] Logistic regression trained using stochastic gradient descent.

- [ ] Linear regression trained using stochastic gradient descent.

5. Which of the following statements about map-reduce are true? Check all that apply.

1 point

- [ ] If you have just 1 computer, but your computer has multiple CPUs or multiple cores, then map-reduce might be a viable way to parallelize your learning algorithm.

- [x] When using map-reduce with gradient descent, we usually use a single machine that accumulates the gradients from each of the map-reduce machines, in order to compute the parameter update for that iteration.

- [x] In order to parallelize a learning algorithm using map-reduce, the first step is to figure out how to express the main work done by the algorithm as computing sums of functions of training examples.

- [ ] Running map-reduce over N computers requires that we split the training set into N^2 pieces.

- [ ] I, **Hassan Rasheed**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.
Learn more about Coursera's Honor Code