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

2. Which of the following statements about stochastic gradient

descent are true? Check all that apply.

1 point

- [ ] In order to make sure stochastic gradient descent is converging, we typically compute $J_{\text{train}}(\theta)$ after each iteration (and plot it) in order to make sure that the cost function is generally decreasing.
- [x] Before running stochastic gradient descent, you should randomly shuffle (reorder) the training set.
- [x] If you have a huge training set, then stochastic gradient descent may be much faster than batch gradient descent.
- [ ] One of the advantages of stochastic gradient descent is that it uses parallelization and thus runs much faster than batch gradient descent.

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

1 point

- [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.
- [ ] 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.
- [x] 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.

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] Logistic regression trained using batch gradient descent.
- [ ] Linear regression trained using stochastic gradient descent.
- [ ] Logistic regression trained using stochastic gradient descent.
- [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).

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

1 point

- [x] Because of network latency and other overhead associated with map-reduce, if we run map-reduce using N computers, we might get less than an N-fold speedup compared to using 1 computer.
- [ ] Linear regression and logistic regression can be parallelized using map-reduce, but not neural network training.
- [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] If you have only 1 computer with 1 computing core, then map-reduce is unlikely to help.

Coursera Honor Code Learn more

- [ ] I, **Vamshi Nenavath**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

Submit Save draft

Like Dislike Report an issue