Quiz

Large Scale Machine Learning

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

- ○ This is not possible with stochastic gradient descent, as it is guaranteed to converge to the optimal parameters θ.
- ○ Try averaging the cost over a smaller number of examples (say 250 examples instead of 500) in the plot.
- ○ Use fewer examples from your training set.
- ◉ Try halving (decreasing) the learning rate α, and see if that causes the cost to now consistently go down; and if not, keep halving it until it does.

2. Which of the following statements about stochastic gradient

1 point

descent are true? Check all that apply.

- [ ] 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.
- [ ] One of the advantages of stochastic gradient descent is that it uses parallelization and thus runs much faster than batch gradient descent.
- [x] If you have a huge training set, then stochastic gradient descent may be much faster than batch gradient descent.

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

1 point

- [x] One of the advantages of online learning is that if the function we're modeling changes over time (such as if we are modeling the probability of users clicking on different URLs, and user tastes/preferences are changing over time), the online learning algorithm will automatically adapt to these changes.
- [ ] Online learning algorithms are most appropriate when we have a fixed training set of size m that we want to train on.
- [x] Online learning algorithms are usually best suited to problems were we have a continuous/non-stop stream of data that we want to learn from.
- [ ] When using online learning, you must save every new training example you get, as you will need to reuse past examples to re-train the model even after you get new training examples in the future.

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.
- [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).
- [ ] 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

- [ ] Linear regression and logistic regression can be parallelized using map-reduce, but not neural network training.
- [x] If you have only 1 computer with 1 computing core, then map-reduce is unlikely to help.
- [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] 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.