✖

Large Scale Machine Learning

5 questions

---

1 point

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?

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

---

1 point

2.
Which of the following statements about stochastic gradient

descent are true? Check all that apply.

- ☐ 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.
- ☐

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.

- [ ] In each iteration of stochastic gradient descent, the algorithm needs to examine/use only one training example.

- [ ] 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.

---

1 point

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

- [ ] 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.

- [ ] 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.

---

1 point

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.

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

---

1 point

5.

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

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

- [x] I 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

Anderson Banihirwe

Submit Quiz