BerkeleyX: CS120x Distributed Machine Learning with Apache Spark

Bookmarks

Bookmark

Model Complexity

(1/1 point)

Imagine we have two regression models, where model A has weights (1.0, 2.0, 1.5) and model B has weights (0.0, 1.0, 0.75). If these two models have the same training error, then under the ridge regression optimization criterion with a positive value for the regularization parameter, which model is more favorable?

- ○ Model A
- ◉ Model B ✔
- ○ The two models are equally favorable

EXPLANATION

The ridge regression optimization has two terms, one related to training error the other related to model complexity. If model A and model B have the same training error, then the model with lower model complexity is more favorable under the ridge regression optimization criterion, and since

Regression and Click-through Rate Prediction

model B has smaller weights it has lower complexity.

Closed-form Solutions

(1/1 point)

Select the machine learning techniques that have closed-form solutions:

- [x] Linear Regression ✔
- [ ] Gradient Descent
- [x] Ridge Regression ✔

✔

Note: Make sure you select all of the correct options—there may be more than one!

EXPLANATION

Linear regression and ridge regression, as explained in the lectures, can be solved exactly. Gradient descent is a general purpose algorithm, which can be used to find the parameter values that minimize a function.

Linear Regression Complexity

(1/1 point)

According to the lecture, which of the following statements about the time and space complexity of linear regression is accurate?:

- [x] O(nd^2 + d^3) computation ✔
- [ ] O(nd + d^2) computation
- [x] O(nd + d^2) storage ✔
- [ ] O(nd^2 + d^3) storage

✔

Note: Make sure you select all of the correct options—there may be more than one!

EXPLANATION

Please refer to the 3 minute mark in the "Distributed Machine Learning: Computation and Storage" video.

Data Growth

(1/1 point)

Which of the following techniques reduce the computational or storage burden when dealing with massive amounts of data?

- [x] Using sparse representations
- [x] Low-rank approximation
- [x] Gradient descent

Note: Make sure you select all of the correct options—there may be more than one!

EXPLANATION

All of these techniques are useful when managing the computational / storage burden of massive datasets.

Gradient Descent

(1/1 point)

Select the true statements about gradient descent:

- [x] Iterative algorithm
- [x] Convergence can be slow

- [ ] Can't be parallelized
- [ ] Always finds global minimum
- [x] Requires communication across nodes ✔

✔

Note: Make sure you select all of the correct options—there may be more than one!

EXPLANATION

Gradient descent is an iterative algorithm that is easily parallelized. Convergence can be slow (many iterations required) and it finds a local minimum (global for convex functions). Gradient descent needs to communicate weights across nodes.

Scalable Algorithms

(1/1 point)
Which of the following communication considerations impact the design of scalable learning algorithms?

- [ ] Memory throughput is only slightly higher than disk throughput
- [ ] Network throughput is substantially higher than disk throughput

- [x] Memory throughput is substantially higher than both disk and network throughput ✔

✔

Note: Make sure you select all of the correct options—there may be more than one!

EXPLANATION

In a given amount of time, we can transfer substantially more data between CPU and memory than we can retrieve from disk or across the network.

Latency

(1/1 point)
Select the true statements about latency:

- [ ] The ratio of memory latency to disk latency is similar in magnitude to the ratio of memory throughput to disk throughput.
- [x] Memory access has lower latency than disk access ✔
- [ ] You can amortize latency by sending updates frequently

✔

Note: Make sure you select all of the correct options—there may be more than one!

EXPLANATION

Latency is much lower for memory access than disk access. According to the lecture, disk latency is more than 10,000 times memory latency, and memory throughput is around 50 times higher than disk throughput. Latency is amortized by sending batch updates (more data sent less frequently).

Some Rights Reserved

© edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.

POWERED BY OPENedX

Download on the App Store

ANDROID APP ON Google play