BerkeleyX: CS120x Distributed Machine Learning with Apache Spark

Bookmarks

- Week 1 - Course Overview, Software Setup, and Machine Learning Basics
- Week 2 - Linear Regression and Distributed Machine Learning Principles
- **Week 3 - Logistic Regression and Click-through Rate Prediction**

Lecture 3: Logistic Regression and Click-through Rate Prediction
Quiz due Aug 08, 2016 at 23:00 UTC

Lab3 - Click-through Rate Prediction Pipeline
Lab due Aug 08, 2016 at 23:00 UTC

Lab3 Quiz

Week 3 - Logistic Regression and Click-through Rate Prediction > Lecture 3: Logistic Regression and Click-through Rate Prediction > Review Quiz

Bookmark

CTR Modeling

(1/1 point)

Why is CTR modeling crucial for online advertising?

☑ Amount advertisers pay is often based on the effectiveness of the ad (e.g., obtaining a click or conversion) ✓

☑ Publishers want to maximize the money they make hosting ads (hence want to host ads with high CTR) ✓

☑ 3rd party matchmakers need to make good matches to stay in business ✓

✓

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

EXPLANATION

All of the above are true. Online advertising generates over $40B in revenue per year, and CTR modeling is important to all participants in the industry.

Quiz due Aug 08, 2016 at 23:00 UTC

Loss Functions

(1/1 point)
What is the purpose of a loss function?

☑ It’s a way to penalize a model for incorrect predictions ✓

☑ It precisely defines the optimization problem to be solved for a particular learning model ✓

☐ It creates new features for use in the model

✓

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

EXPLANATION

Loss functions define how to penalize incorrect predictions. The optimization problems associated with various linear classifiers are defined as minimizing the loss on training points (sometime along with a regularization term).

Convex Loss Functions

(1/1 point)

Which of the following loss functions are convex?

☑ Log-loss ✓

☐ 0 / 1 loss

✓

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

EXPLANATION

Log-loss is convex, which means that we can use gradient descent to find weights that result in a global minimum. 0 / 1 loss is not convex due to its abrubt decision boundary at z = 0, so it is difficult to optimize.

Logistic Regression with Regularization

(1/1 point)

Select the true statements for logistic regression with regularization:

☐ When lambda equals one, it provides the same result as standard logistic regression

☑ Can be framed as minimizing a convex function ✓

☐ Closed-form solution exists

✓

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

EXPLANATION

When lambda is zero, the regularization term is zero and does not affect the model. Logistic regression can be framed as minimizing a convex function but has no closed-form solution.

The Logistic Function

(1/1 point)

The logistic function 1 / (1 + exp^(-z)):

☑ Has a probabilistic interpretation ✓

☐ Approaches 0 for large positive inputs

☑ Returns values between 0 and 1 ✓

✓

EXPLANATION

The logistic function asymptotically approaches 0 as the input approaches negative infinity and 1 as the input approaches positive infinity. Since the results are bounded by 0 and 1, it can be directly interpreted as a probability.

Classification Thresholds

(1/1 point)

When using probabilistic predictions to classify, we should vary the threshold based on the relative harm of false positives relative to false negatives.

● True ✓

○ False

EXPLANATION

True. For example consider the spam filtering example. False positives (legitimate emails erroneously predicted as spam) are likely to cause more harm than false negatives (spam emails that are not identified as spam), as we might miss an important email, while it is easy to delete a spam message. In this case, we could require a higher threshold (probability) that a message is spam before we move it into a spam folder.

Spam Example

(1/1 point)

In the spam example, if we use a threshold of 0 for spam classification, what percentage of emails will be classified as spam?

○ 0%

○ 50%

● 100% ✓

EXPLANATION

A threshold of 0 would result in all emails being classified as spam, as the probability (of spam) for any message would exceed 0 by at least a small margin.

Transforming Categorical Features

(1/1 point)

Representing a categorical feature by a single numeric variable (with a variety of values) can introduce relationships / constraints that were non-existent prior to the transformation.

● True ✓

○ False

EXPLANATION

True. For example, if a survey question had the answers: "Poor, Reasonable, Good, and Excellent" and we assigned those answers the values 1, 2, 3, and 4, respectively, then we would have introduced a relationship where "Excellent" has twice the magnitude of "Reasonable". This relationship did not exist in the prior representation.

Sparse Representations

(1/1 point)
Using a sparse representation of our data can:

☑ Save storage space ✓

☑ Reduce computational costs ✓

✓

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

EXPLANATION

Sparse representations can reduce both space and computational requirements. The relative magnitude of these savings increases with the sparsity of the data.

Feature Collisions

(1/1 point)
With feature hashing it is possible to hash different features to the same bucket.

● True ✓

○ False

EXPLANATION

Feature hashing can be used to reduce the number of features, which means that there will be more features than buckets. All of the features need to map to a bucket, so collisions will occur.

Feature Hashing

(1/1 point)

Feature hashing requires communication of intermediate results across nodes.

○ True

● False ✓

EXPLANATION

False. Hashed features can be computing using a (usually fast) hash function that can be applied independently at the partition level and without communication between nodes.