Week 1

Machine Learning Data Lifecycle in Production

Week 1
Discuss the topic here.

10 threads · Last post 2 days ago

Go to forum

Week 1: Collecting, Labeling and Validating Data

Robert Crowe

This week covers a quick introduction to machine learning production systems. More concretely you will learn about leveraging the TensorFlow Extended (TFX) library to collect, label and validate data to make it production ready.

Learning Objectives

- Describe the differences between ML modeling and a production ML system
- Identify responsible data collection for building a fair production ML system
- Discuss data and concept change and how to address it by annotating new training data with direct labeling and/or human labeling
- Address training data issues by generating dataset statistics and creating, comparing and updating data schemas

∧ Show Less

A conversation with Andrew Ng, Robert Crowe and Laurence Moroney

▶ **Video:** Specialization overview   6 min   Resume

▶ **Video:** Course Overview   2 min

Introduction to Machine Learning Engineering in Production

▶ **Video:** Overview   11 min

▶ **Video:** ML Pipelines   6 min

▤ **Practice Quiz:** Intro to MLEP   10 questions

▤ **Ungraded External Tool:** Join us on Discourse!   1h

Collecting Data

▶ **Video:** Importance of Data   8 min

▶ **Video:** Example Application: Suggesting Runs   8 min

▶ **Video:** Responsible Data: Security, Privacy & Fairness   11 min

▤ **Practice Quiz:** Data Collection   7 questions

Labeling Data

▶ **Video:** Case Study: Degraded Model Performance   9 min

▶ **Video:** Data and Concept Change in Production ML   5 min

▶ **Video:** Process Feedback and Human Labeling   11 min

▤ **Practice Quiz:** Data Labeling   3 questions

Validating Data

▶ **Video:** Detecting Data Issues   7 min

▶ **Video:** TensorFlow Data Validation   6 min

▤ **Practice Quiz:** Issues in Training Data   4 questions

▤ **Reading:** Week 1 Optional References   3 min

▤ **Reading:** How to Download your Notebook   10 min

▤ **Lab:** TFDV Exercise   50 min

Assignment

○ **Reading:** Partial Grading for Assignments 2 min

○ **Programming Assignment:** Data Validation 3h Due Dec 20, 1:59 AM CST

○ **Reading:** Partial Grading for Assignments 2 min

○ **Programming Assignment:** Data Validation 3h Due Dec 20, 1:59 AM CST