Search in course   **Search**

Week 3

Introduction to Machine Learning in Production

Week 3

Discuss the topic here.

6 threads · Last post 2 months ago

Go to forum

Week 3: Data Definition and Baseline

This week is all about working with different data types and ensuring label consistency for classification problems. This leads to establishing a performance baseline for your model and discussing strategies to improve it given your time and resources constraints.

∧ Less

Learning Objectives

- List the questions you need to answer in the process of data definition.
- Compare and contrast the types of data problems you need to solve for structured vs. unstructured and big vs. small data.
- Explain why label consistency is important and how you can improve it
- Explain why beating human level performance is not always indicative of success of an ML model.
- Make a case for improving human level performance rather than beating it.
- Identify how much training data you should gather given time and resource constraints.
- Describe the key steps in a data pipeline.
- Compare and contrast the proof of concept vs. production phases on an ML project.
- Explain the importance of keeping track of data provenance and lineage.

∧ Show Less

Define Data and Establish Baseline

▶ **Video:** Why is data definition hard? 4 min   **Resume**

▶ **Video:** More label ambiguity examples 9 min

▶ **Video:** Major types of data problems 11 min

▶ **Video:** Small data and label consistency 8 min

▶ **Video:** Improving label consistency 9 min

▶ **Video:** Human level performance (HLP) 10 min

▶ **Video:** Raising HLP 9 min

Label and Organize Data

▶ **Video:** Obtaining data 12 min

▶ **Video:** Data pipeline 5 min

▶ **Video:** Meta-data, data provenance and lineage 9 min

▶ **Video:** Balanced train/dev/test splits 4 min

▤ **Quiz:** Data Stage of the ML Production Lifecycle 10 questions   Due Jan 3, 1:59 AM CST

Scoping (optional)

▶ **Video:** What is scoping? 2 min

▶ **Video:** Scoping process 6 min

▶ **Video:** Diligence on feasibility and value 14 min

▶ **Video:** Diligence on value 7 min

▶ **Video:** Milestones and resourcing 2 min

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

▤ **Practice Quiz:** Scoping (optional) 5 questions

Lecture Notes (Optional)

▱ **Ungraded External Tool:** Lecture Notes W3 5 min

○ Ungraded External Tool: Lecture Notes W3

Ungraded Lab

☐ **Lab:** Data Labeling   45 min

Course Resources

◉ **Reading:** References   5 min

Acknowledgements

◉ **Reading:** Acknowledgments   3 min