



Week 5

Machine Learning Modeling Pipelines in Production

Overview

- Week 1
- Week 2
- Week 3
- Week 4
- Week 5

Grades

Notes

Discussion Forums

Messages

Course Info

Week 5

Discuss the topic here.

Go to forum

1 threads · Last post 6 months ago

Week 5: Interpretability



Robert Crowe

Learn about model interpretability - the key to explaining your model's inner workings to laypeople and expert audiences and how it promotes fairness and helps address regulatory and legal requirements for different use cases.

Learning Objectives

- Deconstruct a model to understand and explain how it is making predictions
- Determine tools to make your models explainable to yourself and others in easy to understand terms
- Provide mechanisms to address model fairness and compliance with regulatory requirements
- Differentiate and compare different methods to extract interpretation of models by introspection or explanation
- Distinguish between multiple available techniques such as: Partial Dependence Plots, SHAP and LIME; to provide insight into how model predictions are actually made
- Use cloud managed service for AI explanations

Show Less

Explainable AI

Video: Explainable AI 6 min

Resume

Reading: Explainable AI 3 min

Practice Quiz: Explainable AI 3 questions

Interpretability

Video: Model Interpretation Methods 9 min

Video: Intrinsically Interpretable Models 10 min

Reading: Interpretability 5 min

Practice Quiz: Interpretability 3 questions

Understanding Model Predictions

Video: Model Agnostic Methods 1 min

Video: Partial Dependence Plots 5 min

Video: Permutation Feature Importance 3 min

Reading: Permutation Feature Importance 2 min

Lab: Permutation Feature Importance 30 min

Video: Shapley Values 7 min

Video: SHapley Additive exPlanations (SHAP) 3 min

Lab: Shapley Values 30 min




Reading: Understanding Model Predictions 2 min

Video: Testing Concept Activation Vectors 3 min


Video: LIME 1 min

Reading: TCAV and LIME 2 min




-  **Video:** AI Explanations 5 min
-  **Reading:** AI Explanations 2 min
-  **Practice Quiz:** Understanding Model Predictions 4 questions



Lecture Notes (Optional)

-  **Ungraded External Tool:** Lecture Notes W5 5 min

Course Resources

-  **Reading:** Course 3 Optional References 10 min

Acknowledgments

-  **Reading:** Acknowledgements 3 min
-  **Reading:** Join our DeepLearning.AI Community! 10 min