

Lecture 15 - Contents

An overview of the main sections in this lecture.

Part 1

Attribution and Visualization
Methods

Part 2

Counterfactuals and Explanation
Interfaces

Part 3

Communicating Uncertainty and
Trade-offs

Hands-on

Interpretability Toolkit

This outline is for guidance. Navigate the slides with the left/right arrow keys.

