

# Explainable AI

## Training and Testing a Neural Network

---

Ryan Benasutti

February 28, 2019

Worcester Polytechnic Institute

# Research Goals

- Experimentally demonstrate problems with bias in training data sets
- Show effective methods to test for trained bias

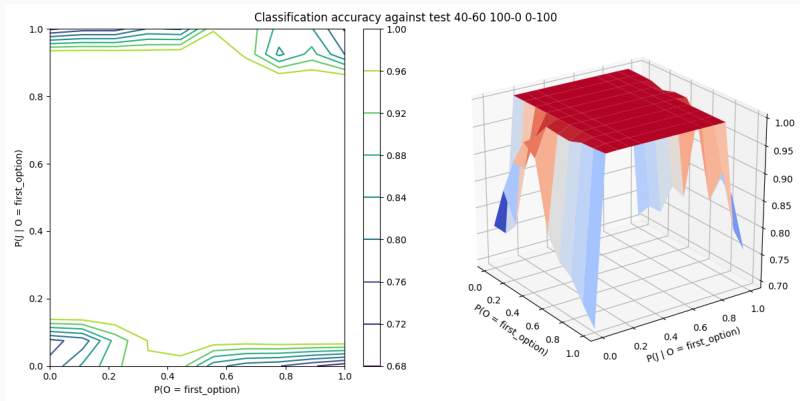
# Modeling the Problem

Dilemma - 2+ Options

Option - 0..10 People

Person - Attributes: Age, Race, Jaywalking, etc.

# Experimental Results



**Figure 1:** Classification Accuracy

# Experimental Results

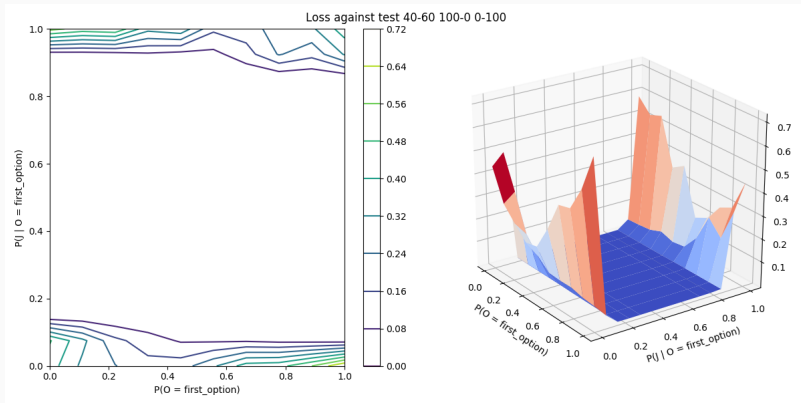
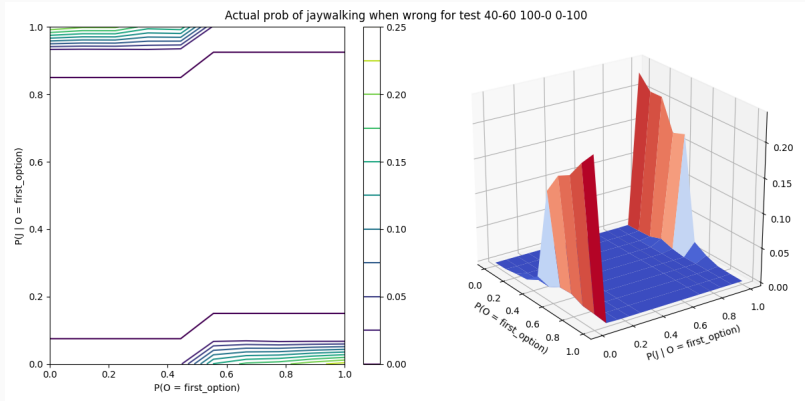


Figure 2: Loss

# Experimental Results



**Figure 3:** Actual Jaywalking Probability