

On the (In)fidelity and Sensitivity for Explanations

Chih-Kuan Yeh, Cheng-Yu Hsieh, Arun Sai Suggala, David Inouye, Pradeep Ravikumar

Problem Definition and Overview

Methods

Experiments