

Introduction to LIME

Local Interpretable Model-agnostic Explanations

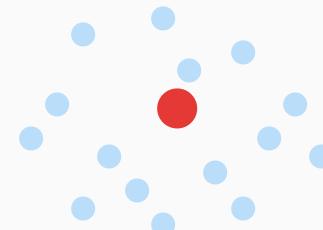
LIME Process (4 Steps)

1 Original Instance



Instance to explain

2 Perturbed Samples



Generate neighbors

3 Local Linear Model



4 Feature Importance



Key Ideas

Core Concepts

- **Local:** Explains one prediction
- **Proximity:** Weight by distance
- **Linear:** Interpretable locally
- **Agnostic:** Any model type

Model-Agnostic

Works with any ML model

Perturbation

Create samples around instance

Weighting

Closer samples matter more

Interpretability

Feature importance shows factors

Python Implementation

Fit weighted model

Extract explanation

lime package:

- lime.lime_tabular
- lime.lime_text
- lime.lime_image