

LIME

Local Interpretable
Model-agnostic Explanations

William Bendinelli

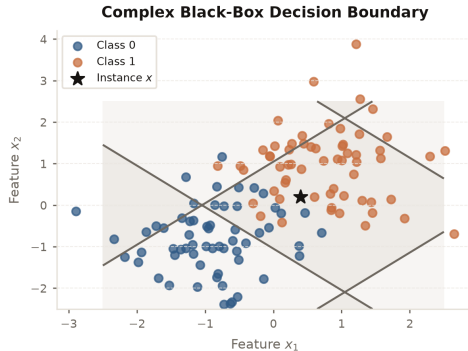
SAPIANS

A prediction is not an explanation.

The model can be confident and still remain opaque.

THE QUESTION

Why this prediction
for this observation?



LIME explains a model locally around one prediction.

It replaces global complexity with a simple approximation in the neighborhood of the instance we want to understand.

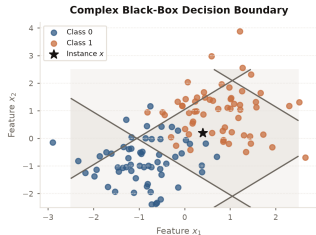
BLACK BOX $f(x)$ **LOCAL****neighborhood****SIMPLE****surrogate****EXPLAIN** β

The equation, read as a design.

$$\xi(x) = \arg \min_{g \in G} \mathcal{L}(f, g, \pi_x) + \Omega(g)$$

- 01 Approximate the black-box model f
- 02 Give nearby samples more importance through π_x
- 03 Penalize complexity through $\Omega(g)$

THE MODEL



One instance, one decision boundary.

THE CODE

```
# black-box classifier
model = SVC(
    kernel='rbf',
    C=2, gamma=2
)
model.fit(X, y)

# decision surface
P = model.predict_proba(grid)
```

THE IDEA

THE IDEA

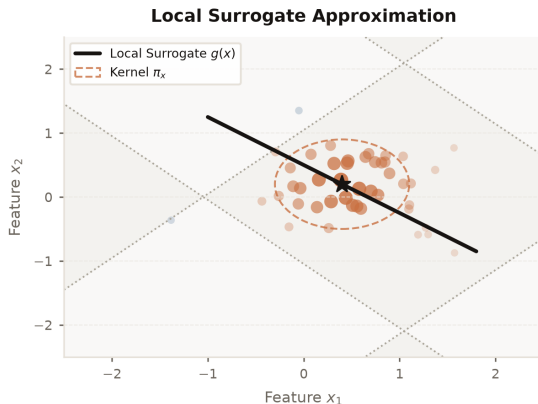
The model can predict.

But prediction is not explanation.

**Explain one prediction without
opening the black box.**

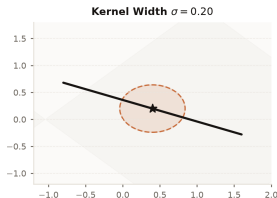
A local surrogate approximates the model near x .

This slide type is for visual evidence. The graph dominates; the explanation stays short.

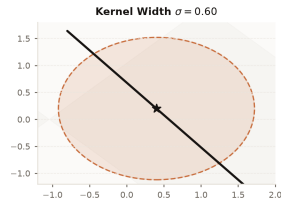


LIME is useful. But not neutral.

- 01 Instability
- 02 Neighborhood
- 03 Surrogate fidelity
- 04 Representation



NARROW kernel = 0.20



WIDE kernel = 0.60

Change the neighborhood — change the local story.

The explanation depends on sampling, weighting and the surrogate's local fidelity.

What does LIME actually explain?

NOT THIS

"How the model works."

THIS

"How the model behaves around this observation."

LOCAL $\neq$ GLOBAL

LIME does not open the black box.

It builds an interpretable approximation of its behavior — locally.

KEY TAKEAWAY

**LOCAL · INTERPRETABLE · MODEL-
AGNOSTIC**

Reusable slide families.

01 COVER

Dark minimal opening & title

02 DEFINITION

Split concept & key tokens

03 EQUATION

Math as editorial design object

04 THREE COLUMNS

Model, Code, and Idea triad

05 EVIDENCE

Graph-led visual proof

06 SYNTHESIS

Not-this/this & dark takeaway