Considerations

Advantages of LIME

Good option to explain models trained on huge data spaces, like, images, text, genomics, etc.

Creates simpler representations, where a group of attributes are switched on or off.

Important ➔ Fidelity

One way of trusting the surrogate is by determining it's fidelity to the black box in the area examined.

High fidelity ➔ we trust the explanations.

Versatility

We can train different models to explain different observations.

➔ Different bag of words

➔ More or less input features

➔ Different depths

The more, the more

The more observations we explain, the more we see why or how the model makes predictions.

If it makes sense for most observations, we are more confident in the model.

Limitations

- We can’t (ultimately) be sure that the surrogate represents well the black box.

- There are a group of possible explanations, how do we know which one is the right one?

Limitations

We need to define various things arbitrarily:

- How do we sample continuous features?
- How big should the super-pixels be?
- How many words in a row should we remove?
- Single bag of words or n-grams?
- How wide the kernel should be?

Limitations

We need to define various things arbitrarily:

- Make the explanations versatile
- We can easily manipulate them (willingly, or, more likely otherwise) to fulfil our expectations of what the explanation should look like.

Train In Data

THANK YOU

www.trainindata.com