Synthetic data

LIME - mechanism

1. Choose the data point to explain.
2. **Generate synthetic data in its proximity.**
3. Obtain the black box predictions for the data from 2.
4. Obtain the distance between synthetic data and original data point.
5. Train a white box with the perturbed data (2) to predict the black box predictions (3), weighted by their locality (4).
6. Interpret the white box.

Text

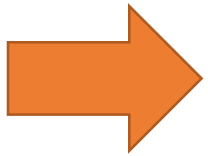
Text

British scientists unveil the ‘world’s first’ laptop powered by light. Mobile computer owners are looking forward for first batch of the laptops.

Generate synthetic data by randomly removing words from the text.

Text

British scientists unveil the ‘world’s first’ laptop powered by light. Mobile computer owners are looking forward for first batch of the laptops.



British ____ unveil the ‘world’s first’ laptop powered by light. Mobile computer owners are looking forward for first batch of the laptops.

British scientists unveil the ‘world’s first’ laptop powered by light. Mobile ____ owners are looking forward for first batch of the laptops.

British scientists unveil the ‘world’s first’ laptop ____ by light. Mobile computer owners are looking forward for first batch of the laptops.

Text

British scientists unveil the ‘world’s first’ laptop powered by light. Mobile computer owners are looking forward for first batch of the laptops.

British unveil the ‘world’s first’ laptop powered by light. Mobile computer owners are looking forward for first batch of the laptops.

British scientists unveil the ‘world’s first’ laptop powered by light. Mobile owners are looking forward for first batch of the laptops.

British scientists unveil the ‘world’s first’ laptop by light. Mobile computer owners are looking forward for first batch of the laptops.

- Use synthetic data to obtain black box predictions.
- Process new text as needed for black box, i.e., embeddings

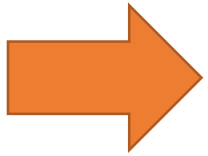
Text

British scientists unveil the ‘world’s first’ laptop powered by light. Mobile computer owners are looking forward for first batch of the laptops.

British unveil the ‘world’s first’ laptop powered by light. Mobile computer owners are looking forward for first batch of the laptops.

British scientists unveil the ‘world’s first’ laptop powered by light. Mobile owners are looking forward for first batch of the laptops.

British scientists unveil the ‘world’s first’ laptop by light. Mobile computer owners are looking forward for first batch of the laptops.



British	scientists	unveil	laptop	powered	computer	owners
1	1	1	1	1	1	1
1	0	1	1	1	1	1
1	1	1	1	1	0	1
1	1	1	1	0	1	1

We create bag of words to train the surrogate.

➔ They are easy to understand.

Images

Images

Divide image into interpretable components (**super-pixels**)

Original Image

Interpretable Components

Images from: https://www.oreilly.com/content/introduction-to-local-interpretable-model-agnostic-explanations-lime/

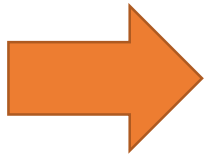
Images

Original Image

Switch-off 1 or more of those components at random.

Images from: https://www.oreilly.com/content/introduction-to-local-interpretable-model-agnostic-explanations-lime/

Images



- Use synthetic data to obtain black box predictions.
- Process new images as needed for the black box.

Images from: https://www.oreilly.com/content/introduction-to-local-interpretable-model-agnostic-explanations-lime/

Images

Superpix 1	Superpix 2	Superpix 3	Superpix …	Superpix n
0	1	1	1	0
0	0	0	0	1
1	1	1	1	0

- Each superpixel is 1 feature.
- It can be on or off

➔ It is easy to understand

Images from: https://www.oreilly.com/content/introduction-to-local-interpretable-model-agnostic-explanations-lime/

Train In Data

THANK YOU

www.trainindata.com