L I M E

Local Interpretable Model-agnostic explanations

InputLayer
X

HiddenLayer
H

OutputLayer
y

$\sim$
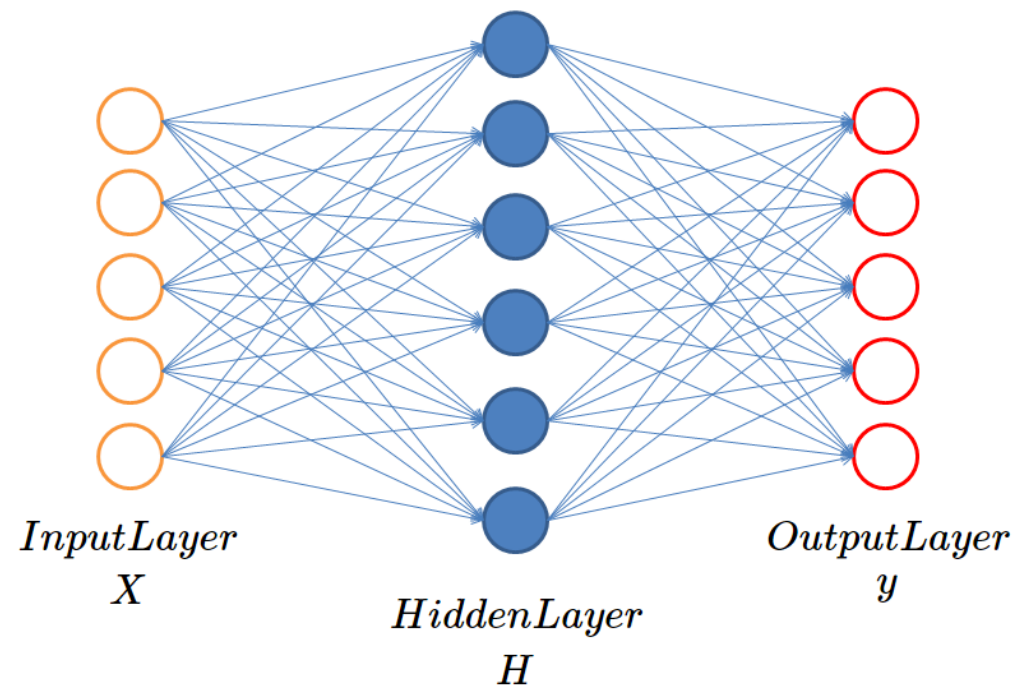
$$\hat{y} = w_1 x_1 + w_2 x_2$$

$$\hat{y} = w_1 x_1 + w_2 x_2$$
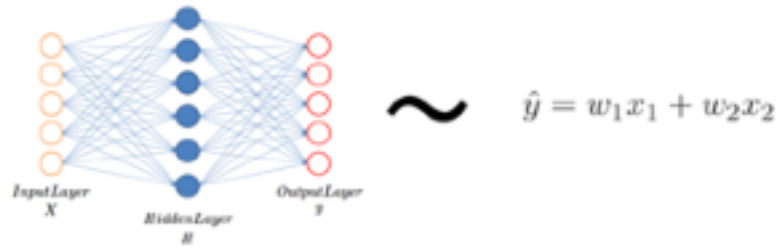
```python
import lime
import lime.lime_tabular
```

```python
explainer = lime.lime_tabular.LimeTabularExplainer(X_train, feature_names=X_train.columns, class_n
ames=[0, 1], discretize_continuous=False)
```

Observation Explanation

```python
exp = explainer.explain_instance(X_train.iloc[1].values, model.predict_proba, num_features=len(X_t
rain.columns), top_labels=1)
```

exp =>