

```
import lime
import lime.lime_tabular
```

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

Observation Explanation

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



Integrated Gradients



Sensitivity

is the attribution sensitive to **relevant features**: inputs differ by a single feature

Axioms

Implementation Invariance

the attributions are always identical of **functionally equivalent** networks

outputs are identical for all inputs