UNIVERSIDADE CATÓLICA PORTUGUESA

BRAGA

Machine Learning

Session 23 - PL

Explainable AI (XAI)

Ciência de Dados Aplicada

2023/2024

XAI with SHAP

- "SHAP (SHapley Additive exPlanations) is a game theoretic approach to explain the output of any machine learning model. It connects optimal credit allocation with local explanations using the classic Shapley values from game theory and their related extensions"

- Installation:
  - pip install shap

  OR
  - conda install -c conda-forge shap

XAI with SHAP

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

[1]:
```
import xgboost

import shap

# get a dataset on income prediction
X, y = shap.datasets.adult()

# train an XGBoost model (but any other model type would also work)
model = xgboost.XGBClassifier()
model.fit(X, y);
```

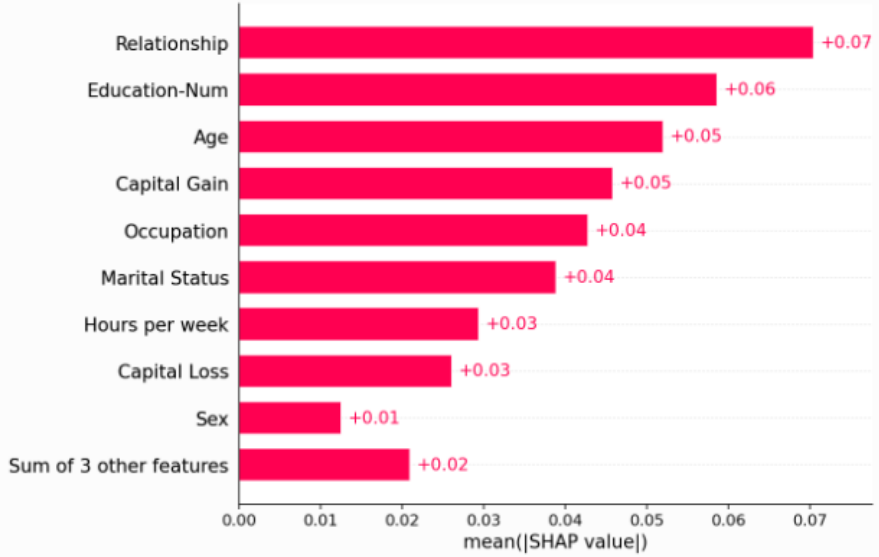
Tabular data with independent (Shapley value) masking

[2]:
```
# build a Permutation explainer and explain the model predictions on the given dataset
explainer = shap.explainers.Permutation(model.predict_proba, X)
shap_values = explainer(X[:100])

# get just the explanations for the positive class
shap_values = shap_values[..., 1]
```

Plot a global summary

[3]:
```
shap.plots.bar(shap_values)
```



Exercises:

- Notebooks on the github repository:
  - Notebook with examples and exercises:
    - notebooks/session_23/exercises.ipynb