.explain_predict(X)

The Objective

- A machine learning model:
  - *model = …* → *fit(X,y)*
- *A prediction:*
  - *model.predict(X)* → *„Iris-virginica“*
- *A prediction that explains itself:*
  - *model.predict_explain(X)* → *„The prediction is Iris-virginica, because …“*

Start with KNeighborsClassifier

- A personalized machine learning model:
  - *model = my_KneighborsClassifier.fit(X,y)*

- *Get prediction that explains itself:*
  - *model.predict_explain(X)* → *Prediction*
    Confidence
    Explanation
    Features_Distribution

model.predict_explain(X)

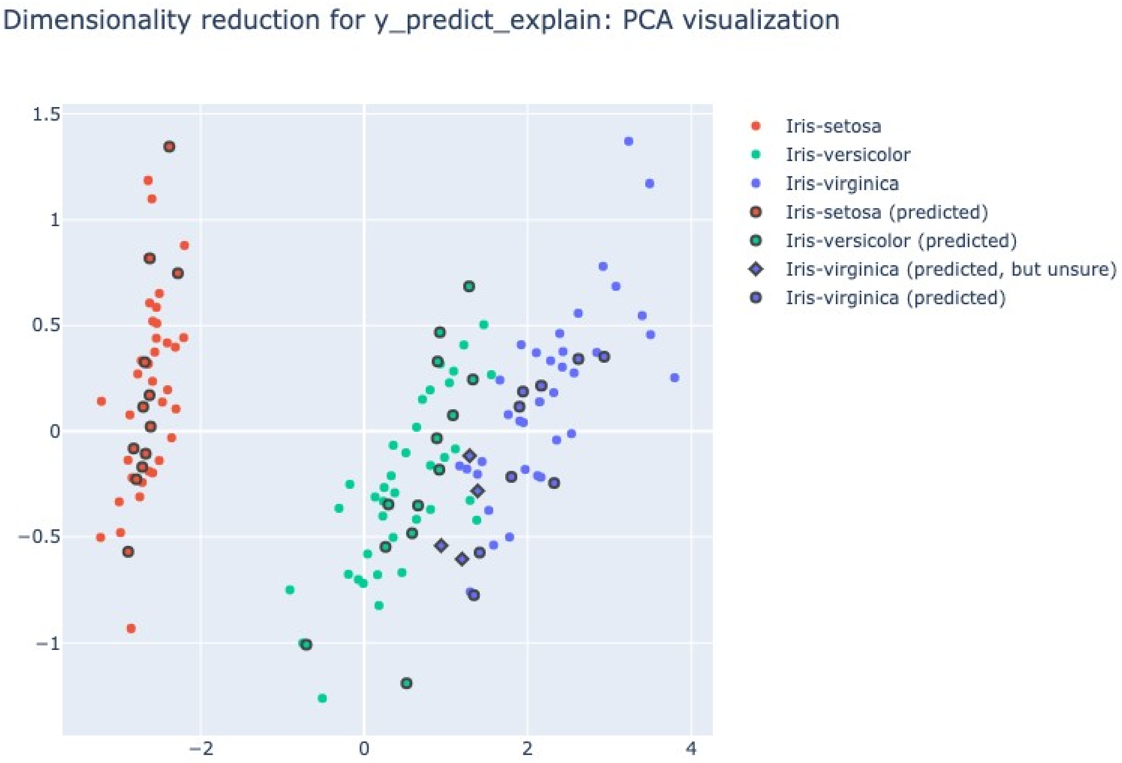
- Prediction: *„Iris-versicolor“*
- *Confidence: False*
- *Explanation:*
  - *„The prediction 'Iris-virginica' is rather unsure: On the one hand the 5 nearest neighbours have diverse target values (2x value 'Iris-versicolor', 3x value 'Iris-virginica'). But on the other hand the nearest neighbour has the same target value too."*

model.predict_explain(X)

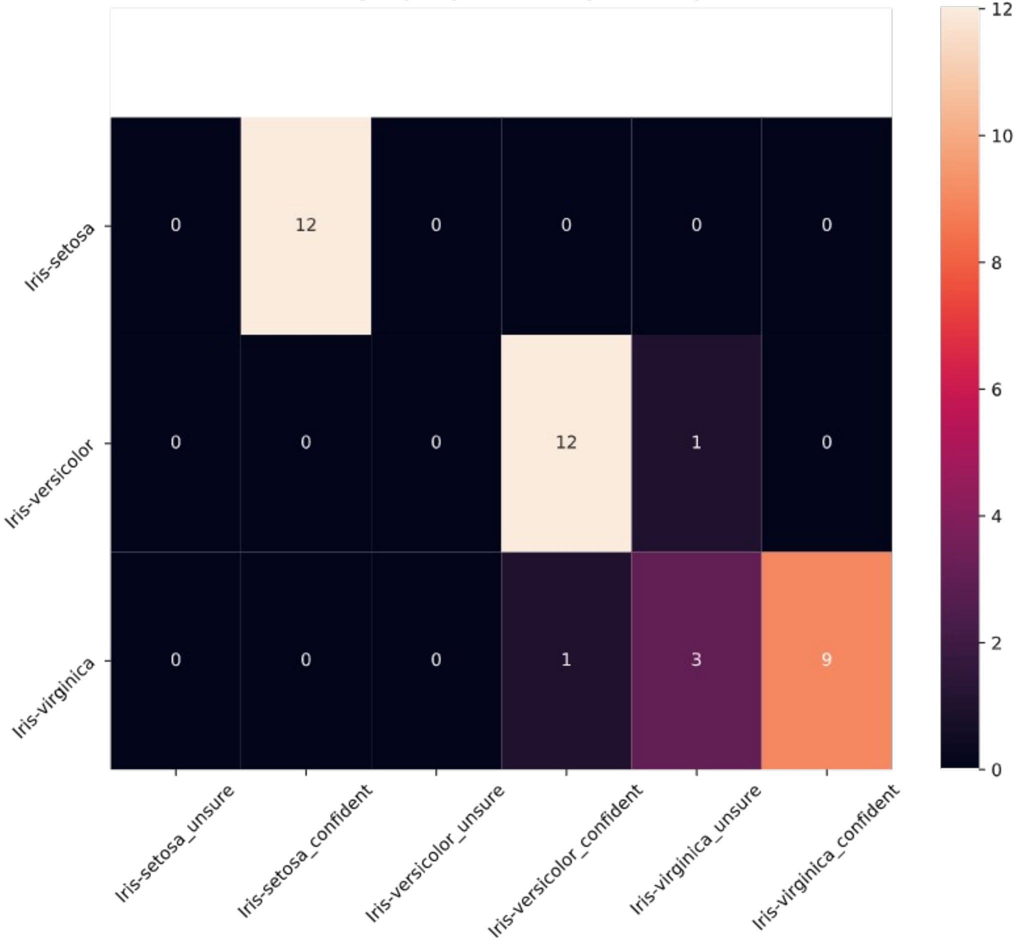
- Features_Distribution:
  - *The features given for predicting the target value are rather far from any other observations already known.*
  - *No feature has the exact same values in the range of the 5 nearest neighbours.*
  - *However, the feature 'sepal_length' differs remarkably ('5.6' vs. '6.0') throughout the inspected 5 nearest neighbours.*
  - *There seems to be an intersection of the target values {'Iris-versicolor', 'Iris-virginica'}."*

Visualization

Interactive



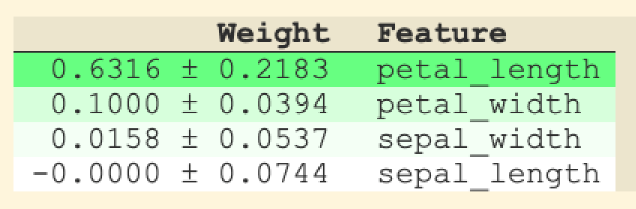
Visualization



	Iris-setosa_unsure	Iris-setosa_confident	Iris-versicolor_unsure	Iris-versicolor_confident	Iris-virginica_unsure	Iris-virginica_confident
Iris-setosa	0	12	0	0	0	0
Iris-versicolor	0	0	0	12	1	0
Iris-virginica	0	0	0	1	3	9

Libraries

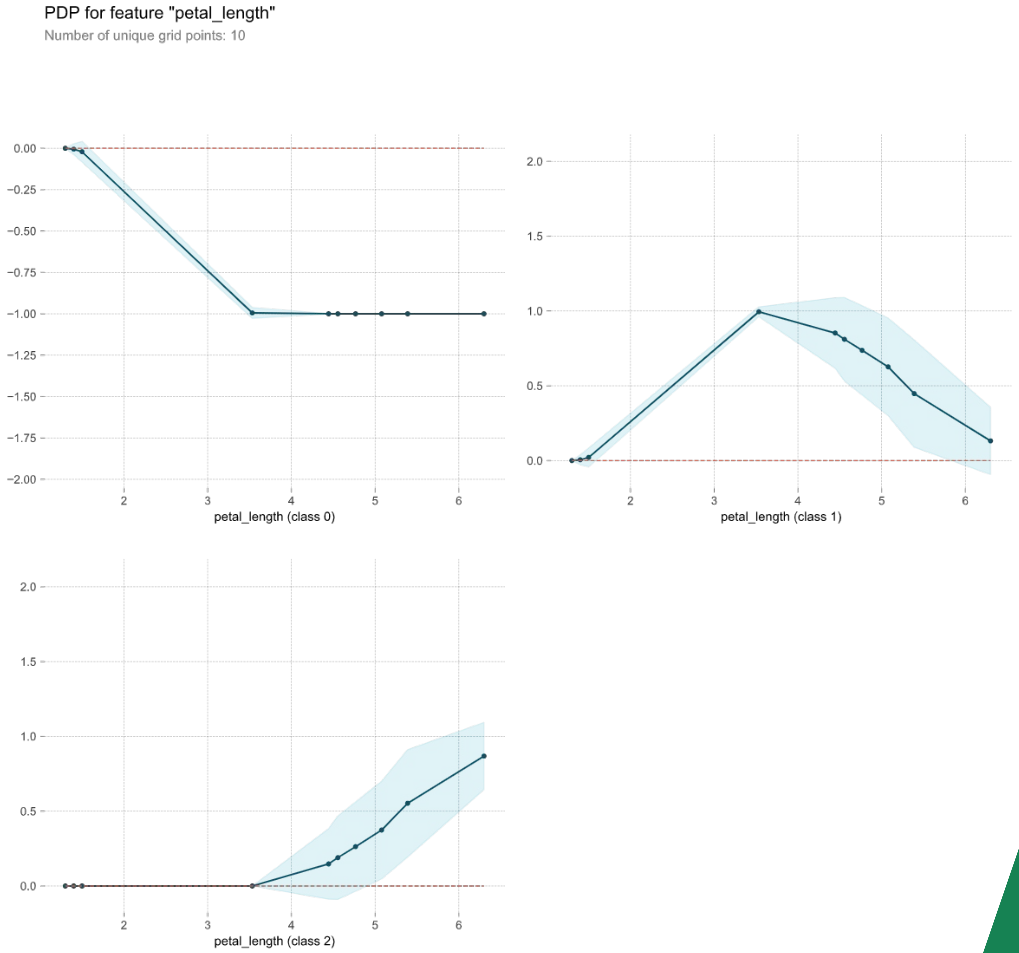
```
from eli5.sklearn import PermutationImportance
```



Weight	Feature
0.6316 ± 0.2183	petal_length
0.1000 ± 0.0394	petal_width
0.0158 ± 0.0537	sepal_width
-0.0000 ± 0.0744	sepal_length

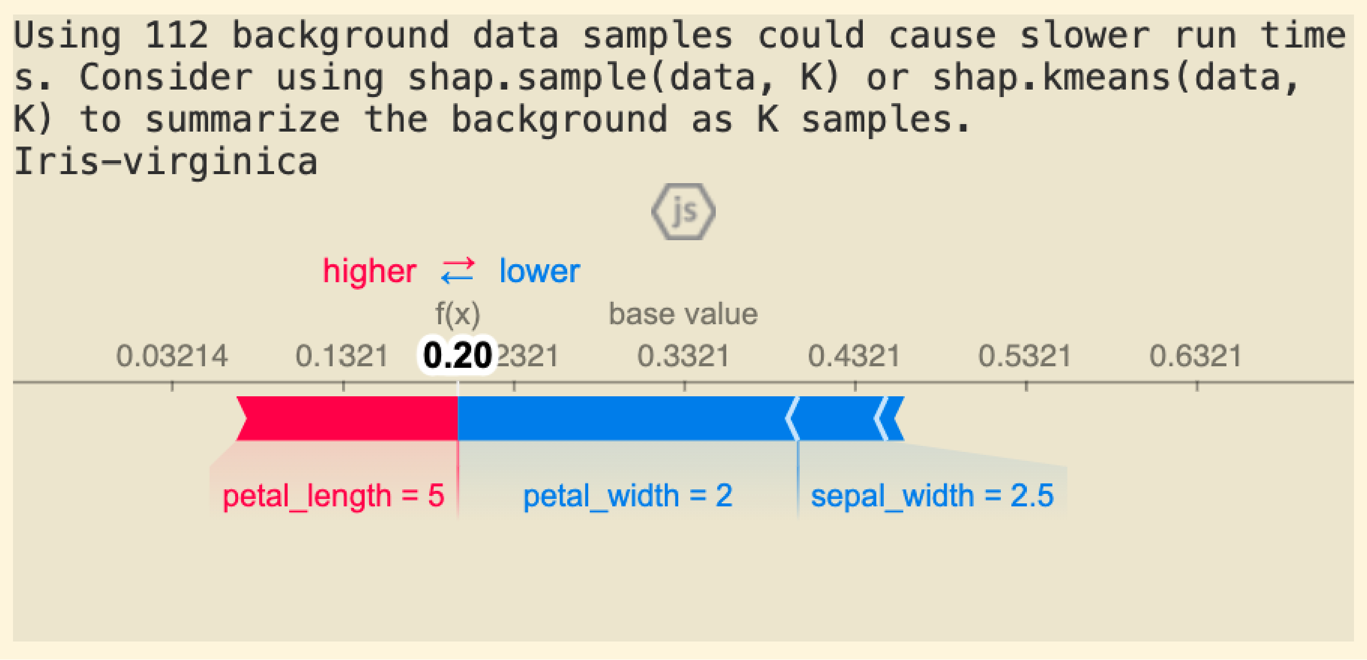
Libraries

```
from pdpbox import pdp
```



Libraries

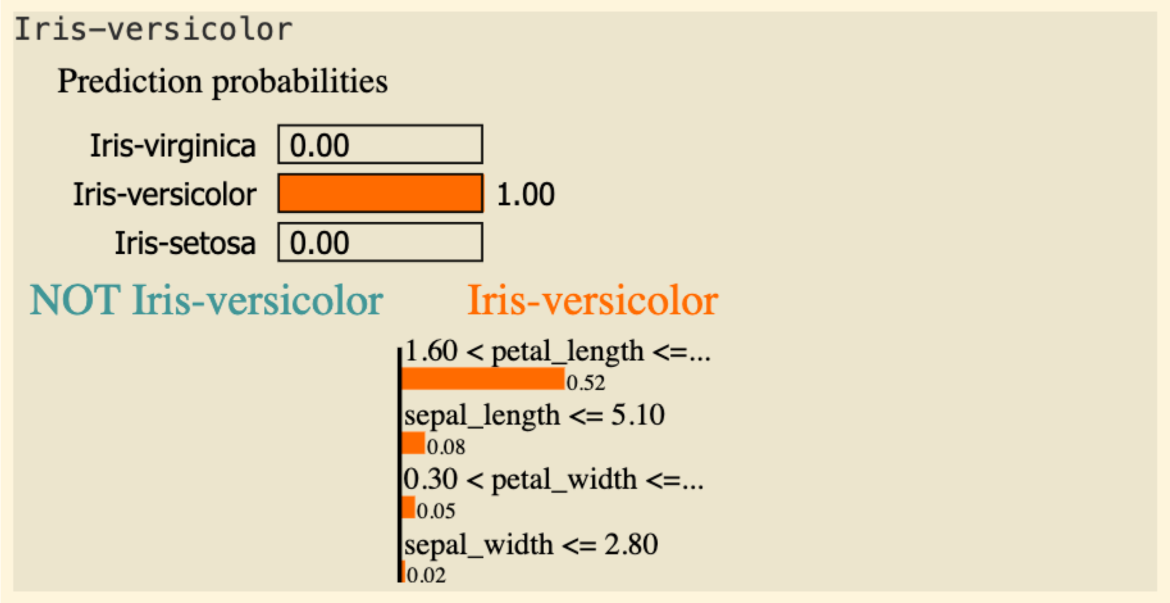
import shap



```
Using 112 background data samples could cause slower run time
s. Consider using shap.sample(data, K) or shap.kmeans(data,
K) to summarize the background as K samples.
Iris-virginica
```

Libraries

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



Näxt steps (?)

- Rollout .predict_explain(X) on RandomForestClassifier
- Dive into explainability models in more detail
- Try it out on a data set
  - Get one from the other groups
  - Get a new one

Merci!