Lab Feature Engineering

Summary

In this lab you will:

- Examine the performance of the Cognito feature engineering algorithm
- Visualize the feature transformations

Refer to the demo videos from this lesson for a step-by-step demonstration of how to complete the lab.

Instructions

1. Explore Feature Engineering for the banknote_auth experiment

    a. Return to the AutoAI Experiment Pipeline leaderboard
    b. Click on the top-ranked pipeline
    c. On the left menu bar, select Feature Transformations
    d. Examine the newly created features
    e. Click on Feature Importance from the left menu bar. Recall that importance is determined by a decision tree-based algorithm.
    f. Notice that the two most important features are curtosis and NewFeature_0, which is a pca transformation of all the features.
        i. PCA (Principal Component Analysis) is a dimensionality reduction technique. Recall that there are 4 attributes in this dataset: variance, skewness, curtosis and entropy. NewFeature_0 is the first principal component resulting from the dimensionality reduction procedure. For more on PCA, see here: https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html

2. Inspect the Cognito transformations via the notebook

    a. Open the original notebook
    b. In the Compose Pipeline cell, find the code block with the comment `# composing steps for cognito Pipeline`
    c. Notice the two transforms that are applied are `pca` and the `sigmoid` function.
    d. Look up the documentation for the `TAM, FS1, TA1` and `FS1` classes from autoai_libs here: https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/autoai-lib-python.html

3. Visualize the transformation NewFeature_0

    a. Under 5. Preprocess Data, following the cell with comment `# extract X and y`, insert a new cell and type the following code:

```
!pip install yellowbrick
from yellowbrick.features import PCA

visualizer = PCA(scale=True, proj_features=True)
```

```
visualizer.fit_transform(df_X, df_y)
visualizer.show()
```

b. PC1 is the feature NewFeature_0 used in the model

4. Save this version of the notebook

5. Optional: Repeat these steps for the parkinsons_updrs experiment.

   a. Notice the importance of NewFeature_5 and NewFeature_0.
      i. Visualize NewFeature_0 in the same way as above.
      ii. Visualize the transformation for NewFeature_5 with the following code:

```
import matplotlib.pyplot as plt

def plot_feature(x, y, x_lab, y_lab):
    fig = plt.figure()
    ax = fig.add_axes([0,0,1,1])
    ax.scatter(x,y)
    ax.set_xlabel(x_lab)
    ax.set_ylabel(y_lab)
    plt.show()

plot_feature(df['age'], df['total_UPDRS'],  "age",
"total_UPDRS")

plot_feature(np.tan(df['age']), df['total_UPDRS'] ,
"tan(age)", "total_UPDRS")
```