Multiclass classification: Takeaways ⤴

Syntax

- Return a DataFrame containing binary columns:

```python
dummy_df = pd.get_dummies(cars["cylinders"])
```

- Concatenate DataFrames:

```python
cars = pd.concat([cars, dummy_df], axis=1)
```

- Return a Series containing the index of the maximum value:

```python
predicted_origins=testing_probs.idxmax(axis=1)
```

Concepts

- In the instance where 2 values are just 2 different labels, it is safer to turn the discrete values into categorical variables.

- Dummy variables are for columns that represent categorical values.

- A problem is a multiclass classification problem when there are three or more categories or classes. There are existing multiclassification techniques that can be categorized into the following:

  - **Transformation to Binary:** Reduce the problem to multiple binary classification problems.

  - **Extension from Binary:** Extend existing binary classifiers to solve multi-class classification problems.

  - **Hierarchical classification:** Divide the output into a tree where each parent node is divided into multiple child nodes and the process is continued until each child node represents only one class.

- The one-versus-all method, which is a transformation to binary technique, is a technique where we choose a single category as the Positive case and group the rest of the categories as the False case.

Resources

- [Documentation for idxmax()](#)
- [Multiclass Classification](#)