AI Bootcamp

Machine Learning Pipeline and Mini-Project

Module 12 Day 3

Class Objectives

By the end of class, you will be able to:

1. Build a machine learning pipeline.
2. Apply supervised learning regression algorithms in a mini-project.

Activity:

Regression Warm Up

In this activity, you will warm up your regression skills by applying adjusted R-squared using linear regression on rent pricing data.

Suggested Time:
15 Minutes

Time’s up!

Let’s review

Questions?

Instructor **Demonstration**

Machine Learning Workflows

The machine learning **workflow** defines how each phase of the project unfolds and in which order.

Why Use a Machine Learning Workflow?

1. Much of the code used for machine learning is repeated for every project.

2. Every machine learning engineer will develop a “flow” that works best for them.

3. While a time investment in the beginning, building a set of utilities that can streamline your workflow in the future can be incredibly helpful later.

What **steps** are commonly repeated across the models we've built so far?

What steps are commonly repeated across the models we've built so far?

01 Loading data and cleaning it by removing nulls and running **`get_dummies`** on the dataset.

02 Splitting data into training and testing sets.

03 Finding and selecting the best features for use in a model.

04 Training and scoring the model.

05 Trying multiple model setups and comparing them.

Which **situations** might benefit from being able to train multiple models at the same time using a precoded workflow?

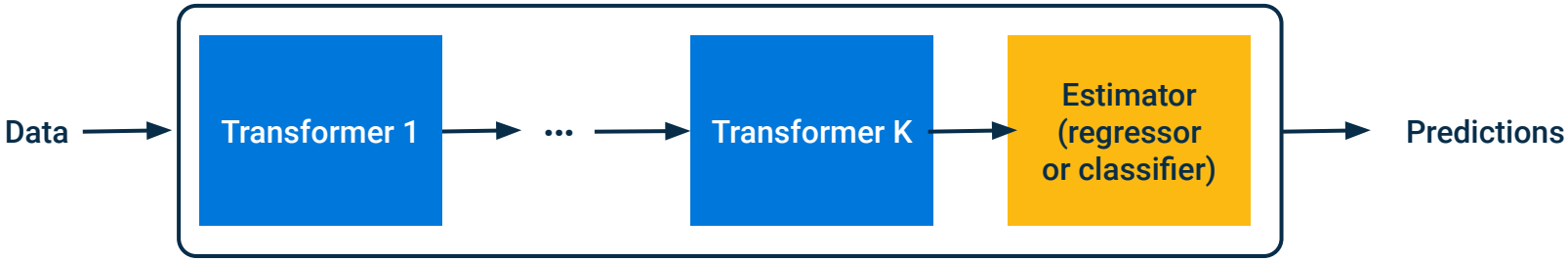
Which situations might benefit from being able to train multiple models at the same time using a precoded workflow?

01 Training a model to forecast sales from multiple stores. Instead, train one model per store and combine the predictions.

02 Training a model to predict the payroll of a company. Instead, train one model per department and combine the predictions.

03 Training individual models to predict weather in multiple states. Instead, use a workflow to train the models.

Understanding Pipeline in Machine Learning with Scikit-learn




```
from sklearn.pipeline import Pipeline
# creating pipeline and fitting it on data
pipe = Pipeline([('transformer', StandardScaler(with_mean=False)),
                 ('estimator', LinearRegression())
                ])
```

The transformers we've used so far are like **OneHotEncoder** and **StandardScaler**.

Machine Learning Workflows

Set up the list of tuples, which are one or more preprocessing transformers, followed by the estimator (regression or classification model):

```
steps = [
    ("One hot encode", OneHotEncoder(handle_unknown="ignore")),
    ("Scale", StandardScaler(with_mean=False)),
    ("Linear Regression", LinearRegression())
]
```

Parse the list of tuples as a parameter:

```
pipeline = Pipeline(steps)
```

To use the pipeline to fit the model, call the `fit` method with the X data you want to fit.

```
pipeline.fit(X_train, y_train)
```

Make Predictions

```
# Use the pipeline to make predictions
y_pred = pipeline.predict(X_test)
```

Instructor **Demonstration**

Machine Learning Pipelines

Instructor **Demonstration**

Intro to Mini-Project

Activity:

Mini-Project Part 1 - Explore Datasets

In this activity, you will work in groups to explore and select a dataset that you will use to plan, build, train, and evaluate a linear regression model, combining the steps into a machine learning workflow.

Suggested Time:
10 Minutes

Time’s up!

Let’s review

Activity:

Mini-Project Part 2 - Data Preprocessing and Model Training

In this activity, you will work together to combine your skills with data preprocessing to prepare your dataset and train the model.

Suggested Time:
30 Minutes

Break

15 mins

Activity:

Mini-Project Part 3 - Write ML Workflow Application

In this activity, you will have a single function that can be imported and used to train and evaluate a model on the dataset.

Suggested Time:
20 Minutes

Time’s up!

Let’s review

Activity:

Mini-Project Part 4 - Test Application

In this activity, you will put your ML workflow application to the test.

Suggested Time:
25 Minutes

Time’s up!

Let’s review

Activity:

Discuss and Share Mini-Projects

In this activity, you will present your models to the class.

Suggested Time:
25 Minutes

Questions?

Review the Class Objectives

In this lesson, you learned how to:

1. Build a machine learning pipeline.
2. Apply supervised learning regression algorithms in a mini-project.

Questions?

The End