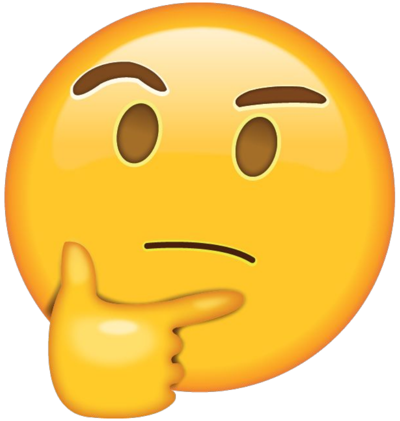
Object Oriented Programming

@cambridgespark

CAMBRIDGE SPARK

Wait, why?

(It’s a Data Science course …)



CAMBRIDGE SPARK

Object Oriented Programming

For Data Scientists

@cambridgespark

CAMBRIDGE SPARK

Let’s pause for a minute...

- What do we have so far?
- What do we want to achieve?

CAMBRIDGE SPARK

What do we have so far?

A large collection of algorithms that:

- take some parameters

What do we have so far?

A large collection of algorithms that:

- take some parameters
- need to be trained on some data

CAMBRIDGE SPARK

What do we have so far?

A large collection of algorithms that:

- take some parameters
- need to be trained on some data
- once trained, can be used to generate predictions

CAMBRIDGE SPARK

What do we want to achieve?

As Data Scientists, we want to be able to **pick an algorithm**, **train** it on our data, **test** it and maybe **deploy** it to a production environment

CAMBRIDGE SPARK

What do we want to achieve?

As Data Scientists, we want to be able to **pick an algorithm**, **train** it on our data, **test** it and maybe **deploy** it to a production environment

We need:

- A library with all the algorithms
- A consistent way to **initialise** them with the right parameters
- A consistent way to **train** them on our data
- A way to keep the **internal state** in memory
- A consistent way to generate **predictions**

CAMBRIDGE SPARK

What do we want to achieve?

As Data Scientists, we want to be able to **pick an algorithm**, **train** it on our data, **test** it and maybe **deploy** it to a production environment

We need:

- A library with all the algorithms
- A consistent way to **initialise** them with the right parameters
- A consistent way to **train** them on our data
- A way to keep the **internal state** in memory
- A consistent way to generate **predictions**
- A way to produce **maintainable** and **structured code** that larger team can work on and use

That can be achieved with **Object Oriented Programming**

CAMBRIDGE SPARK

Definition

“**Object-oriented programming** (OOP) is a programming paradigm based on the concept of "**objects**", which may contain data, in the form of fields, often known as **attributes**; and code, in the form of procedures, often known as **methods**.”

CAMBRIDGE SPARK

Definition

“**Object-oriented programming** (OOP) is a programming paradigm based on the concept of "**objects**", which may contain data, in the form of fields, often known as **attributes**; and code, in the form of procedures, often known as **methods**.”

- Classes are used to define the structure of an object:
  - What attributes and methods it will have
- Objects can be instantiated from a class, with a set of parameters
  - That will initialise the internal state
  - This is called an instance of the class
- Methods can change the internal state of our object

An example

```
# DataFrame is a class, df is an instance of this class
df = pd.DataFrame({"age": [20, 21], "name": ["Alice", "Bob"]})

# We can access attributes on df
df.columns

# We can call methods on df
df.drop("age", axis=1, inplace=True)

# That will drop the column age, and modify the internal state of
our object
```

CAMBRIDGE SPARK

What do we want to achieve?

As Data Scientists, we want to be able to **pick an algorithm**, **train** it on our data, **test** it and maybe **deploy** it to a production environment

We need:

- A library with all the algorithms
- A consistent way to **initialise** them with the right parameters
- A consistent way to **train** them on our data
- A way to keep the **internal state** in memory
- A consistent way to generate **predictions**
- A way to produce **maintainable** and **structured code** that larger team can work on and use

That can be achieved with **Object Oriented Programming**

CAMBRIDGE SPARK

What do we want to achieve?

As Data Scientists, we want to be able to **pick an algorithm**, **train** it on our data, **test** it and maybe **deploy** it to a production environment

We need:

- A library with all the algorithms -- **Create a class per algorithm**
- A consistent way to **initialise** them with the right parameters
- A consistent way to **train** them on our data
- A way to keep the **internal state** in memory
- A consistent way to generate **predictions**
- A way to produce **maintainable** and **structured code** that larger team can work on and use

That can be achieved with **Object Oriented Programming**

CAMBRIDGE SPARK

What do we want to achieve?

As Data Scientists, we want to be able to **pick an algorithm**, **train** it on our data, **test** it and maybe **deploy** it to a production environment

We need:

- A library with all the algorithms -- **Create a class per algorithm**
- A consistent way to **initialise** them with the right parameters -- **Add attributes**
- A consistent way to **train** them on our data
- A way to keep the **internal state** in memory -- **Add attributes**
- A consistent way to generate **predictions**
- A way to produce **maintainable** and **structured code** that larger team can work on and use

That can be achieved with **Object Oriented Programming**

CAMBRIDGE SPARK

What do we want to achieve?

As Data Scientists, we want to be able to **pick an algorithm**, **train** it on our data, **test** it and maybe **deploy** it to a production environment

We need:

- A library with all the algorithms -- **Create a class per algorithm**
- A consistent way to **initialise** them with the right parameters -- **Add attributes**
- A consistent way to **train** them on our data -- **Use methods**
- A way to keep the **internal state** in memory -- **Add attributes**
- A consistent way to generate **predictions** -- **Use methods**
- A way to produce **maintainable** and **structured code** that larger team can work on and use

That can be achieved with **Object Oriented Programming**

Build you own classes

```
class Model:

    def __init__(self, depth):
        self.depth = depth
        self.mean = None
```

This is a parameter to the constructor. It could be named anything you want that is relevant to your domain or algorithm. In this case we called it `depth`.

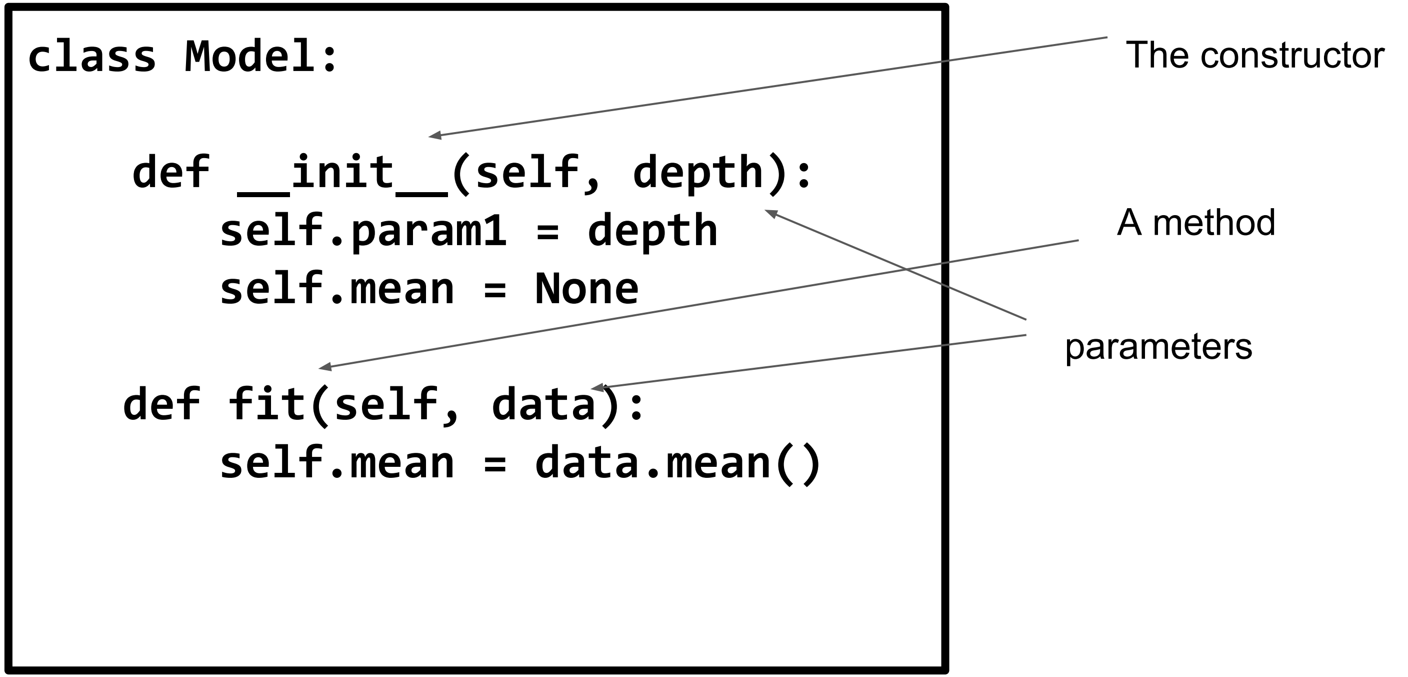
CAMBRIDGE SPARK

Build you own classes

```
my_model = Model(10)

my_model.depth

my_model.mean

type(my_model)
```

CAMBRIDGE SPARK

Build you own classes



```python
class Model:

    def __init__(self, depth):
        self.param1 = depth
        self.mean = None

    def fit(self, data):
        self.mean = data.mean()
```

Build you own classes

```
my_model = Model(10)

my_model.fit([1, 2, 3])

my_model.mean
```

CAMBRIDGE SPARK

Build you own classes

```
class Model:

    def __init__(self, depth):
        self.depth = depth
        self.mean = None

    def fit(self, X_train, y_train):
        self.mean = y_train.mean()

    def predict(self, X_test):
        return [self.mean] * len(data)
```

Build you own classes

```
my_model = Model(10)

my_model.fit(["a", "b", "c"], [1, 2, 3])

my_model.predict(["d", "b", "a"])
```

CAMBRIDGE SPARK

Scikit-Learn

You do not need to implement all algorithms from scratch...

CAMBRIDGE SPARK

Scikit-Learn

You do not need to implement all algorithms from scratch...

The Scikit-Learn community did it for you

- Main algorithms available with a consistent API
- Preprocessing tools
- Metrics
- Model selection
- …

Scikit-Learn

You do not need to implement all algorithms from scratch...

The Scikit-Learn community did it for you

- Main algorithms available with a consistent API
- Preprocessing tools
- Metrics
- Model selection
- ...

Currently 1000+ contributors on Github:
https://github.com/scikit-learn/scikit-learn

scikit learn

CAMBRIDGE SPARK

But we’ll do it anyway

- It’s important to understand how models are implemented
- You will need to implement your own custom models
    - In fact you will need it to submit on K.A.T.E. as well

We’ll implement **not_sklearn**, our own library that follows the scikit-learn conventions!

K.A.T.E. Machine Learning Projects

Demo Time

CAMBRIDGE SPARK