Lab 02 - The basics of serialisation in Python

Welcome to Lab 02 here we will look very quickly at the basics of using pickle. We have looked at this in the slides and we have seen what Pickle is used for. We will be using Pickle to take the model we create and serialise a version of it to our local file system.

Let's begin by looking at a super simple example from the Pickle wiki.

Step 1

If you have not used pickle before then you will need to import the library.

```
python -m pip install pickle
```

Once installed you can execute the following lines of python in jupyter.

1. Open a new Jupyter notebook on your local machine.
2. Create a new notebook - Python
3. Add the following code to the first cell

```
import pickle
```

4. Run previous cell and add the following code to the next cell

```
favorite_color = { "lion": "yellow", "kitty": "red" }
print(favorite_color)
```

5. Run previous cell and add the following code to the next cell

```
pickle.dump( favorite_color, open( "save.p", "wb" ) )
```

6. Run previous cell and add the following code to the next cell
7. Add the following code to the next cell

```
favorite_color_pickle = pickle.load( open( "save.p", "rb" ) )
print(favorite_color_pickle)
```

8. Run previous cell
9. Open save.p you will see the serialised version of that object.

Step 2

Lab 02 - The basics of serialisation in Python

In this step we will pickle the output of our model and then score using it.

10. Amend your python model code to reflect the code below. Either in a notebook or as a PY file.

```
import _pickle as pickle
import numpy as np
from sklearn import datasets, linear_model

PickleModelPath = './regression.pkl'

diabetes = datasets.load_diabetes()
diabetes_X = diabetes.data[:, np.newaxis, 2]
diabetes_X_train = diabetes_X[:-20]
diabetes_X_test = diabetes_X[-20:]
diabetes_y_train = diabetes.target[:-20]
diabetes_y_test = diabetes.target[-20:]
regr = linear_model.LinearRegression()
regr.fit(diabetes_X_train, diabetes_y_train)

diabetes_y_pred = regr.predict(diabetes_X_test)
with open(PickleModelPath, 'wb') as f:
        pickle.dump(regr, f)
```

11. Run the python script to serialise the model.
12. Navigate to where regression.pkl was created, you can see the serialised version of the model.

Step 3

13. Create a new python notebook, but this time do not import the same libraries as before.
14. Run the following.

```
import _pickle as pickle

PickleModelPath = './regression.pkl'

var1 = 10

with open(PickleModelPath, 'rb') as k:
        PickleModel = pickle.load(k)
Answer = PickleModel.predict(var1)

print(Answer)
```

We are now able to score our model without the need to import SKlearn.

We can also store this model in source control.

Lab 02 - The basics of serialisation in Python