Machine Learning Project Documentation

Deployment

Overview

Deployment is a key step in an organization gaining operational value from machine learning. The simplest way to deploy a machine learning model is to create a Web Microservice like an REST API. But to make it more acessable,such that even Non-technical people can interact with the model,we'll also be creating a Web Interface for it.
Below diagram describes the architecture of the system after deployment,it describes how both the developer and a user exchange data with the ML model.

Model Serialization

I used serialization with Pickle, Pickle is a Python module used for serializing and deserializing Python objects. It converts Python objects into a byte stream that can be saved to disk and later reconstructed into the original object.
To serialize a trained model, pickle converts the model object into a format that can be written to a file.

```
import pickle

pickle.dump(dect_model, open("dect_model.pkl", "wb"))
```

Detailed

Serving a serialized machine learning model for making predictions typically involves setting up a prediction service that can load the model and use it to make predictions on incoming data. The choice of deployment platform, whether cloud services or on-premises solutions, depends on several factors such as scalability, cost, security, and specific project requirements. Here's how this process generally works:

- The prediction service starts by deserializing the model (using pickle.load in Python, for instance) to reconstruct the trained model object.
- This model is then kept in memory, ready to make predictions.
- A common approach is to set up a RESTful API using web frameworks like Flask or Django for Python.
- The API defines endpoints that receive data, pass it to the model for prediction, and return the prediction results.

- The service listens for incoming prediction requests, which could be HTTP requests containing new data points.
- Upon receiving a request, the service processes the data (ensuring it's in the correct format for the model), makes a prediction, and sends back the result.
- Cloud Services: Platforms like Netlify offer robust, scalable, and flexible hosting solutions.