Planning Capstones

1 Capstone: Build a ML model of stock prediction

Solution: Use Quandl free data to train a model of stock prediction. The training data will be from Quandl, such as the free WIKI data of a company. The model may be a linear regression model or a Keras model.

The project focuses data processing and model training. This is a data-focused project. Estimated work effort as below:

(1) Data collection & processing: 40%

(2) Model prototype and designing: 20%

(3) Training and deployment: 40%

2 Capstone: build an object detector

Solution: use specific image data to train a model of object detection. After training, the model (object detector) should be able to recognize a set of objects from images or from live video.

Data source: use kaggle image data such as the Oxford-IIIT Pet Dataset

The project focuses image data processing and object detection. This is a data-focused project. Estimated work effort as below:

(1) Data collection & processing: 20%

Get in hand

Grasp major methods of data collection. Collect image data for training and testing an object detection model. Clean and process data to remove possible bad bias. Make all images similar size. Find a good way to labeling data, etc. Build an automated data pipeline

(2) Model prototype and designing: 40%

Design a CNN model with several layers. For hidden layers, output filters, kernel size, activation functions will be adjusted during model training

(3) Training and deployment: 40%

Use existing deployment method to train and test the model and get a model with best training result.

3 Capstone: Netflix movie recommendation model

Solution: use existing data to train a model, in order to get a movie recommendation model

The project focuses model and architecture. Estimated work effort as below:

(1) Data collection & processing: 30%

Use existing and processed data.

(2) Model prototype and designing: 45%

Try different models and make ensemble models to enhance performance. Design different layers in a model. Adjust kernel size and activation functions, etc.

(3) Training and deployment: 25%

Build automation to train/test data. Design own engineering architecture.