CE/CZ 4042: Neural Networks and Deep Learning

Programming Assignment

Part A: Classification problem

- DNN to classify the GTZAN dataset: http://marsyas.info/downloads/datasets.html
- 1000 audio tracks, spanning 30 seconds each.

- The dataset has been pre-processed and **57 features** has been extracted: **features_30_sec.csv.**

- There are **10 different genres** to classify:

blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae and rock.

- Begin with **start_1a.ipynb.**

Part A:

1. DNN with one hidden layer (16 ReLU units), SGD with ‘adam’ optimizer. Dropout at $p = 0.3$. Divide the dataset into 70:30 train and test. Use early-stopping
2. Use 3-fold CV to determine the optimal batch size from { 1, 4, 8, 16, 32, 64}. Report time-taken (use Callbacks).
3. Use 3-fold CV to determine the optimal number of hidden-layer neurons from {8, 16, 32, 64}
4. Implement DNN with two hidden layers.
5. Study the effect of Dropouts

Part B: Regression problem

- The aim is to predicting housing prices in Singapore from related features.
- **Numeric** features: dist_to_nearest_stn, dist_to_dhoby, degree_centrality, eigenvector_centrality, remaining_lease_years, floor_area_sqm
- **Categorical** features: month, flat_model_type, storey_range
- **HDB_price_prediction.csv**
- Start with **start_1b.ipynb**

Part B:

1. DNN with one hidden layer. Divide the dataset into **Train data**: up to year 2020; **Test data**: for year 2021; one-hot encoding for categorical variables:
   - **dataframe_to_dataset** (csv to keras.dataset)
   - **encode_numerical_feature** (normalization)
   - **encode_categorical_feature** (one-hot-encoding)

2. Use **an embedding layer** to encode categorical variables:

tf.keras.layers.Embedding()

3. Use **Recursive Feature Elimination (RFE)** to remove irrelevant features:

Remove irrelevant features one-by-one

Notes

- Based on the **report** (in pdf) and accuracy of **codes (**in .zip file**)**
- 45 for Part A + 45 for Part B + 10 for presentation
- Delayed submissions will be **penalized** for 5 marks for each day up to 3 days
- Absolutely **NO copying, duplicating,** or **plagiarism**.
- Post your queries on Discussion Board
- Approach TAs **Charlene** and **Yihao** for help