AER850 Introduction to Machine Learning

Assignment 3 Rev 2

Marks. 10 Due: March 26th 2023, 11:59pm

- This is an individual assignment.
- Submission instructions: Make a jupyter notebook file into your local TMU drive. Add kdkumar@torontomu.ca and "Acunsolo@torontomu.ca as the collaborators. **Name the file as 'Last_First_AER 850_Assignment 3. **
- Also download the file as an .ipynb file WITH OUTPUTS and the file name (*Last_First_AER 850 Assignment 3)* and upload it to D2L.
- Submit the **assignment report along with your codes** on D2L by the due date.

Bearing Fault Multi-Classification Problem

Bearing is one of the most widely used machine element in rotating machines; its failure can cause serious breakdown of a machine. Figure 1 shows a typical rolling element bearing, which consists of the cage, ball, inner race and outer race.

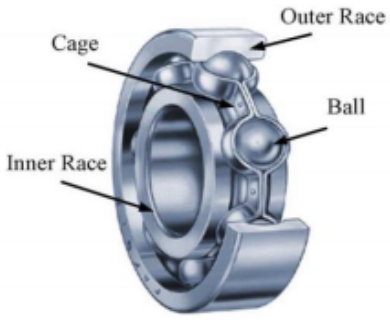

Figure 1. A Typical Rolling Element Bearing (https://arxiv.org/abs/1901.08247)

You are given a bearing dataset (https://engineering.case.edu/bearingdatacenter/download-data-file) which contains vibration data of normal and fault bearings. The data was recorded for motor loads of 0 to 3 hp (motor speed of 1797 to 1720 RPM) using accelerometers at the drive end (DE) and fan end (FE) and the data is stored as Matlab files. The sampling rate is 12 kHz/ 48 kHz and each Matlab file contains between ~120k to ~240k sample points. It is a **multiclass classification** problem, where in the input to the AI model will be the vibration signal data at DE while the output will be the type of defects:

- 0 : Normal (N),
- ~~1 : Fault at Ball (B),~~

- 2 : Fault at Inner Raceway (IR),
- ~~3 : Fault at Outer Raceway (OR),~~

Build an artificial neural network (ANN) or feed forward neural network (FFNN) or convolutional neural network (CNN) classifier to make sense of the input data and classify the type of bearing defects in the bearing's inner race. You may utilize internet sources and past lab codes as a baseline to get you started. I guarantee that chatGPT will be just as confused as you, so don't even bother trying to cheat with that.

Table 1 shows which files to use from the website as baseline data, and Table 2 shows which data sets to use for the inner race. Use 50% of the complete baseline data for training, and 50% for testing. Within the training half, use 80% for training and 20% for validation. Compare your classification results to those acquired using a 70%/30% split.

For your convenience, the DE data is now simplified and available in an excel file posted alongside this update. Use the Baseline DE, 0.007 DE and 0.021 DE for now. A sample script for opening excel files into a pandas dataframe is also provided for VS Code. When running in Colab, you will need to upload the excel file to each runtime and modify the directory as necessary.

Report on the classification results on the training and test data sets.

Table 1. Normal Baseline Data.

Fault Diameter	Motor Load (hp)	Motor Speed (rpm)	Data Set
Normal (no fault)	0	1797	Normal_0
	1	1772	Normal_1
	2	1750	Normal_2
	3	1730	Normal_3

Table 2. Inner Race Fault Data.

Fault Diameter	Motor Load (hp)	Data Set
0.007"	0	IR007_0
	1	IR007_1
	2	IR007_2
	3	IR007_3
0.021"	0	IR021_0
	1	IR021_1
	2	IR021_2
	3	IR021_3

```
import pandas as pd
import os
#simple script to import an excel file into a pandas dataframe
#in terminal, make sure you use pip install openpyxl
cwd = os.getcwd()
file_name = "AER850 A3 Modified Data.xlsx"
files = os.listdir(cwd)
file_path = os.path.join(cwd, file_name)

print(files)
df1 = pd.read_excel(file_path, sheet_name="Baseline DE")
df2 = pd.read_excel(file_path, sheet_name="0.007 DE")
df3 = pd.read_excel(file_path, sheet_name="0.021 DE")

print("Sheet1:")
print(df1.head(5))
print("\nSheet2:")
print(df2.head(5))
print("\nSheet3:")
print(df3.head(5))
```

```
PS D:\Downloads\Python files>  d:; cd 'd:\Downloads\Python files'; & 'C:\Users\AMCun\AppData\Local\Microsoft\WindowsAp
thon\debugpy\adapter/../..\debugpy\launcher' '53027' '--' 'D:\Downloads\Python files\AER850Marking\A3\Code Starter.py'
['.vscode', 'AER850 A3 Modified Data.xlsx', 'AER850Marking', 'bearing_data.csv', 'Chess Pieces', 'DS 8015 Files']
Sheet1:
   Baseline 0HP DE X97  Baseline 1HP DE X98  Baseline 2HP DE X99  Baseline 3HP DE
0             0.053197             0.145667             0.064254         0.014603
1             0.088662             0.097796             0.063002         0.054449
2             0.099718             0.054856            -0.004381         0.107646
3             0.058621             0.036982            -0.035882         0.133722
4            -0.004590             0.054445            -0.023991         0.112652

Sheet2:
   12K 0.007" 0HP  12K 0.007" 1HP  12K 0.007" 2HP  12K 0.007" 3HP
0       -0.083004       -0.277602       -0.093238        0.222699
1       -0.195734       -0.044345        0.187288        0.093238
2        0.233419        0.117603        0.217663       -0.146516
3        0.103958       -0.145055        0.070172        0.177217
4       -0.181115       -0.111430        0.100385        0.248526

Sheet3:
   12K 0.007" 0HP  12K 0.021" 1HP  12K 0.021" 2HP  12K 0.021" 3HP
0        1.189431        0.171369       -0.402027       -0.206293
1       -0.177866        0.117765        0.548219       -0.007310
2       -0.774816       -0.097055        0.931565        0.222536
3        0.501518        0.009746       -0.218881       -0.005685
4        0.993697        0.060913       -1.079788        0.031675
```