

COMP 3223 Coursework

Dr Jo Grundy, Dr. Sri Dasmahapatra

November 3, 2022

1 Introduction

Use `scikit-learn` for splitting data up for testing and training, (from `sklearn.model_selection` import `train_test_split`). However, please use your own implementations for the machine learning.

◇ *You should submit:*

1. Report, pdf format, 4 pages long
2. Code, in zip format

2 Regression

Use the function $f(x)$,

$$f(x) = e^{-\frac{1}{2}(x-2)^2} \sin\left(\frac{3\pi}{2}x\right) - e^{-\frac{1}{2}(x+2)^2} \cos(\pi x),$$

to generate $\{(x_n, y_n)\}_{n=1}^N$ pairs, with x_n randomly drawn from $-5 \leq x_n \leq 5$, $y_n = f(x_n)$ and $N = 30$. This is your training set. A test set can be generated using the same mechanism.

◇ *For the report:*

- Fit a range of different orders of polynomial
- Evaluate the models using a test set and a suitable performance metric
- Plot the performance vs. polynomial order
- Describe and evaluate your work

3 Classification

Classify the Iris dataset ¹ using Fishers' linear discriminant analysis. You can load the data thus: `from sklearn.datasets import load_iris`.

◇ *For the report:*

- Describe the model used
- Evaluate the model using a test set and a suitable performance metric
- Fully describe and evaluate your work

¹https://en.wikipedia.org/wiki/Iris_flower_data_set is at <http://archive.ics.uci.edu/ml/datasets/Iris>