

MANE 4962: Machine Learning for Engineering

Homework 6

Points:100

This homework is designed to test your understanding of PCA, SVM, decision trees, random forests, feature importance, and recurrent neural networks. Use a Jupyter notebook to solve individual problems. You can submit a pdf printout of the notebooks. Please include your name and RIN in the PDF. Please keep your notebooks and Python codes organized in your Github repository.

1. (a) **(10 pts.)** Calculate the first two principal components of the Scikit-learn / UCI ML repository Wine Dataset using Scikit's PCA algorithm.
(b) **(10 pts.)** What is the explained variance ratio of the first two principal components?
(c) **(10 pts.)** Use the first two principal components to train a SVM classifier to classify the wines into three classes. Use a 60%-40% split for the train set and the test set.
(d) **(10 pts.)** Compare the accuracy, precision and recall of the SVM classifier with that of a decision tree classifier that classifies the wines in to three classes using the original features. Use the same 60%-40% split for the train set and the test set for the decision tree classifier and use `max_depth=3`.
(e) **(5 pts.)** Plot the decision tree classifier.
2. **(15 pts.)** Calculate the feature importance of the pixels in the images of the CIFAR-10 dataset (combine train and test sets) using a random forest classifier.
3. **(40 pts.)** Import the surface temperature data, `surface_temp.npy`, for a small sphere from LMS. The temperature data is given for 1000 timesteps. Predict the temperatures for the next 10 timesteps using a recurrent neural network.