

Machine Learning



Lab06: Perceptron & Support Vector Machine classifiers

The goal of this lab exercise is to compare the linear perceptron classifier with the support vector machine classifier on a dataset of handwritten digits.

Task 1.

In this task you will load and visualise the NIST dataset containing labelled examples of normalised images of hand-written digits.

- A. Load the dataset from the dataset repository of SciKit learn using the `datasets.load_digits()` function.
- B. Print the description of the dataset. What are the features and what are the labels of this dataset?
- C. Visualise for each of the possible class labels one random feature vector as images using the `plt.imshow()` function.
- D. Split the dataset into a training and a test set.

Task 2.

In this task you will evaluate the performance of the Perceptron and the SVM classifier on the NIST dataset using k-fold cross validation.

- A. Setup a cross-validation procedure that splits the training dataset from task 1 into a training and a validation set.
- B. Within this cross-validation procedure, train a Perceptron classifier on the training set and evaluate its accuracy score on the validation set.
- C. Within the same cross-validation procedure, also train a Support Vector Machine classifier on the training set and evaluate the accuracy score on the validation set for different choices of kernel function and its parameters.

Task 3.

In this task you will identify and visualise ambiguous features that were mis-classified by the best performing classifier.

- A. Select the best performing classifier from task 2 and predict labels for the test set separated in task 1.
- B. Identify the mis-classifications and determine for which class label the classifier performed worst.
- C. Visualise some random mis-classified feature vectors as images.