

# Homework 3

March 2022

## 1 Support vector machines (SVM)

Solve a two-class classification problem for images using SVM. The dataset is sampled from 2 categories (“T-shirts” and “Trousers”) of the Fashion-MNIST dataset.

- **Dataset Information**

The **training set** contains 12000 images stored in `X_train_sampled.npy`; the labels of the training samples are stored in `y_train_sampled.npy`. The **testing set** contains 2000 images stored in `X_test_sampled.npy`; the labels of the test samples are stored in `y_test_sampled.npy`. Each image has the size of  $28 \times 28$ .

- **Experiment Procedure**

1. Extract the Histogram Of Gradient (HoG) features of the image (page 14 of lecture 2). For a sample  $x \in \mathbf{R}^{28 \times 28}$ , the extracted HoG feature vector is denoted as  $h_x \in \mathbf{R}^{1764}$ .
2. Using the extracted HoG feature vector  $h_x$ , try to use different SVM classifiers for classification. You can implement SVM using the scikit-learn library. You need to implement three SVM classifiers with outliers (page 13 of lecture 2):
  - (a) Linear SVM;
  - (b) RBF SVM;
  - (c) SVM with arbitrary kernel function, e.g. Polynomial SVM.  
You need to find the appropriate value for each parameter in the kernel function, e.g., the appropriate value of  $C$  (the coefficient of  $\xi$ ) (page 13 of lecture 2).

• **Please provide the following results:**

1. Classification accuracy of SVM on the test set.
2. For Linear SVM, find the support vectors involved in the calculation of the parameter  $w$  (page 7 of lecture 2)
  - (a) How many support vectors are involved in the calculation of the parameter  $w$ ? How many of them are positive samples and negative samples?
  - (b) You need to find the images of the support vectors, and the weights corresponding to each support vector. These training samples satisfy  $-y_i(w^T X_i + b) + 1 = 0$  (page 13 of lecture 2). Please include at most 30 support vectors with the largest weights  $y_i \alpha_i$ .