

Machine Learning Lab 6

Principle Component Analysis

In this lab, you will use Principle Component Analysis (PCA) for dimension reduction in classification tasks. As for the dataset, we still use the MNIST which we used in our previous labs. You should finish the following tasks.

1. Use `pca` function from `sklearn.decomposition` to reduce the original feature to 20%, 50%, and 80% amount of information. Then test the reduced feature using KNN to evaluate the classification performance. You should use KNN from sklearn. For more details, see <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html> . See the following as an example:

```
from sklearn.decomposition import PCA
pca = PCA(n_components=1)
X_small_trans = pca.fit_transform(X_small)
X_test_trans = pca.transform(X_test)
```

Then, then you should use `X_small_trans` for the training and `X_test_trans` for testing.

2. Use `pca` function from `sklearn.decomposition` to reconstruct the original feature from the reduced features at different percentage obtained from task 1. And then select few sample images (for example, first 10 images from the training data) and compare the reconstructed version with the original version by plotting them side by side. See the following as an example for reconstruction:

```
from sklearn.decomposition import PCA
X_small_trains_reconst =
pca.inverse_transform(X_small_trans)
```