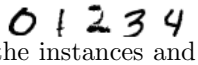
The objective of this lab is for you to explore the behavior of PCA and LDA in Matlab and apply them to some datasets. The TA will first demonstrate the results that PCA and LDA give on the MNIST dataset. Then, you will replicate those results, and further explore other datasets.

We provide you with the following:

- The scripts `lab05_pca.m` and `lab05_lda.m` set up the problem (MNIST) and plot various figures. The computation of PCA and LDA is inlined in the script.

I Datasets

You will use the MNIST dataset of handwritten digits  . PCA will need the instances $\mathbf{x} \in \mathbb{R}^D$ (where $D = 784$), while LDA will need both the instances and their labels $y_n \in \{0, \dots, 9\}$. You will need to plot instances as grayscale images of 28×28, as seen in previous labs; and "eigendigits" as color images of 28×28. You will also need to plot reduced-dimension instances $\mathbf{z}_n \in \mathbb{R}^L$ (where L is 1D, 2D or 3D) as scatterplots; color them differently for each class (even if the class information was not used for training), so we can tell them apart.

Additionally, you will apply PCA and LDA to:

- The rotated-7 MNIST dataset  . Each digit '7' should be considered as a class containing all its rotated versions. Ignore the "skeleton" data in the file, just use the images and the class labels.
- Other datasets, for example from the UCI repository, or a dataset of face images (several are available in the Internet).

Important: it is instructive to test PCA first with toy examples for which you know the true solution ahead of time (e.g. generate points along a line in 3D and add noise to them, then reduce to 1D or 2D with PCA). Once you understand this, try more difficult datasets.

II Using PCA

See the file `lab05_pca.m`. Assume a matrix $\mathbf{X}$ of $D \times N$ (instances = columns).

- To estimate the PCA parameters, we compute the mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ of the data, and then compute the eigendecomposition of the covariance matrix $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$ and set $\mathbf{W} = \mathbf{U}_{1:L}$.
- With this, we can now:
  - Project a point $\mathbf{x} \in \mathbb{R}^D$ onto the L principal component subspace (where $1 \leq L \leq D$). This is given by the PCA projection mapping $\mathbf{z} = \mathbf{F}(\mathbf{x}) = \mathbf{W}^T(\mathbf{x} - \boldsymbol{\mu})$.
  - Reconstruct a vector $\mathbf{z} \in \mathbb{R}^L$ into the original, data space. This is given by the reconstruction mapping $\mathbf{x}' = \mathbf{f}(\mathbf{z}) = \mathbf{W}\mathbf{z} + \boldsymbol{\mu}$.
- We can verify that the covariance matrix in the projected space (that is, cov $\{\mathbf{z}_1, \dots, \mathbf{z}_N\}$) equals $\mathbf{W}^T\boldsymbol{\Sigma}\mathbf{W}$, that it is diagonal, and that the sum of its diagonal elements equals $\lambda_1 + \cdots + \lambda_L$.
- We plot the following figures:
  1. The eigenvalues $\lambda_1, \dots, \lambda_D$ and the proportion of explained variance $\frac{\lambda_1 + \cdots + \lambda_L}{\lambda_1 + \cdots + \lambda_D} \in [0, 1]$ as a function of the number of dimensions used L.
  2. The mean $\boldsymbol{\mu}$, as a grayscale image.
  3. The MNIST dataset projected onto 2D. We use different colors/markers for different digit classes, so we can recognize them.
  4. The MNIST dataset projected onto 3D, colored as in the 2D plot.
  5. The eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_L \in \mathbb{R}^D$, as color images ("eigendigits").
  6. A vector $\mathbf{x}$ and its reconstruction $\mathbf{x}' = \mathbf{W}(\mathbf{W}^T(\mathbf{x} - \boldsymbol{\mu})) + \boldsymbol{\mu}$, both as grayscale images.

7. Vectors of the form $\boldsymbol{\mu} \pm \alpha \mathbf{u}_l$ for $\alpha > 0$ (where $1 \leq l \leq D$), as grayscale images. This shows what the lth principal component subspace corresponds to in data space. It is equivalent to reconstructing vectors $\mathbf{z} \in \mathbb{R}^L$ that move along the lth PC axis.

Then, explore PCA in different settings:

- Compute PCA on only the digits 1s, then visualize it and reconstruct digits (1s, 2s, etc.). The projection on the first two PCs shows a clear structure, what does it correspond to? Why does the mean $\boldsymbol{\mu}$ look the way it does?
- Compute PCA on the entire MNIST dataset (all digits), then visualize it and reconstruct digits.
- See the end of file `lab05_pca.m` for further suggestions.

III Using LDA

See the file `lab05_lda.m`. Assume a matrix $\mathbf{X}$ of $D \times N$ (instances = columns) and a vector $\mathbf{y}$ of $1 \times N$ (class labels in $1, \ldots, K$).

- To estimate the LDA parameters, we compute the within-class and between-class scatter matrices $\mathbf{S}_W$ and $\mathbf{S}_B$. It is convenient to add a small number to the diagonal of $\mathbf{S}_W$ (e.g. $10^{-10} \operatorname{tr}(\mathbf{S}_W)/D$) to make $\mathbf{S}_W$ be full rank. Then we compute the eigendecomposition of $\mathbf{S}_W^{-1}\mathbf{S}_B = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$ and set $\mathbf{W} = \mathbf{U}_{1:L}$.
- With this, we can now project a point $\mathbf{x} \in \mathbb{R}^D$ onto the LDA subspace of dimension L (where $1 \leq L \leq K-1$). This is given by the LDA projection mapping $\mathbf{z} = \mathbf{F}(\mathbf{x}) = \mathbf{W}^T\mathbf{x}$.
- We plot the following figures:
  1. The eigenvalues $\lambda_1, \ldots, \lambda_D$ and the proportion of explained variance $\frac{\lambda_1+\cdots+\lambda_L}{\lambda_1+\cdots+\lambda_D} \in [0,1]$ as a function of the number of dimensions used L.
  2. The mean of each class $\boldsymbol{\mu}_k$, as a grayscale image.
  3. The MNIST dataset projected onto 2D. We use different colors/markers for different digit classes, so we can recognize them.
  4. The MNIST dataset projected onto 3D, colored as in the 2D plot.
  5. The eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_L \in \mathbb{R}^D$, as color images ("Fisherdigits").

Questions to consider:

- Explore the algorithm in different settings (see the PCA section), and with different numbers of classes (for MNIST: different digits).
- How does the result of LDA differ from that of PCA? In particular, observe the 2D projections and the eigendigits and Fisherdigits.
- How many eigenvalues are nonzero in LDA (and how many in PCA)? Why?
  Remember that LDA applies if $\mathbf{S}_W$ is invertible and $L \leq K-1$.