Walking through the projects

Principal Component Analysis(PCA) for dimensionality reduction and visualization

Smita Bhattacharya
M.Sc. in Data Science & Artificial Intelligence
Saarland University

29th February, 2024
Saarland, Germany

PCA for dimensionality reduction and visualization(Cont..)

- **Goal:**
  - Project high dimensional data to a specific low dimensional space.

- **Dataset:** CIFAR 10- an inbuilt torchvision dataset

- **Labels:** 10 different classes
  - ('plane', 'car', 'bird', 'cat', 'deer', 'dog', 'frog', 'horse', 'ship', 'truck)

- **Input Dimension:** 3072(3* 32* 32)

- **Output dimension:** 50

PCA for dimensionality reduction and visualization(Cont..)

Steps:

1. Transformation is done by centering and normalizing the images
2. Transformed matrix(data matrix) is of shape [No of images, 3072]
3. Covariance matrix is computed of the data matrix by multiplying it with the transposed form
4. Eigenvectors and eigenvalues of the covariance matrix is computed by torch.eig function
5. Eigenvectors are sorted by decreasing eigenvalues
6. Top target_dim eigenvectors is selected to get the encoding matrix of shape [target_dim(50), dimensions]
7. Data matrix is multiplied with the encoding matrix to project the data into the low dimensional space

PCA for dimensionality reduction and visualization(Cont..)

- PCA is often used for visualization purposes.
  - Visually exploring the data can become challenging when we have more than 3 features.
  - A very useful tool when dealing with data related problems
- Input : 3072 dimensions
- Output: 2 dimensions
- Plotting the data
  - For each data point, plot the first principle component on x axis
  - and the second principle component on y axis,
  - use different colors for each class.
  - Set corresponding labels: assign label "first principle component" for x and y axis
  - Add legends for each class.