Machine Learning – COMP3032

Tutorial and Lab Practice 6 – Week 7

This lab pactice focuses on the concepts and techniques of unsupervised learning: Dimensionality Reduction.

Tutorial

1. Review the terminology and concepts introduced and algorithems taught in Lecture 6
2. What are the main motivations for reducing a dataset's dimensionality? Are there any drawbacks?
3. What is PCA? What is it used for?
4. In what case would you use Incremental PCA?

Lab Practice

1. Download, open and run the program tut06.py. Read and understand the program.

   For the MNIST example,

   1) Understand each figure plotted
   2) Change the minimum total variance ratio to preserve to 85%, what is the new dimension? Compare the figures with the 95% case
   3) Change the minimum total variance ratio to preserve to 98%, what is the new dimension? Compare the figures with the previous two cases
   4) Add another way to reduce the dimension, based on the variance to preserve, by directly initialising the PCA() object. (using the *n_components* variable)
   5) Suppose you want to reduce the number of dimension to 100, what is the overall preserving variance ratio?