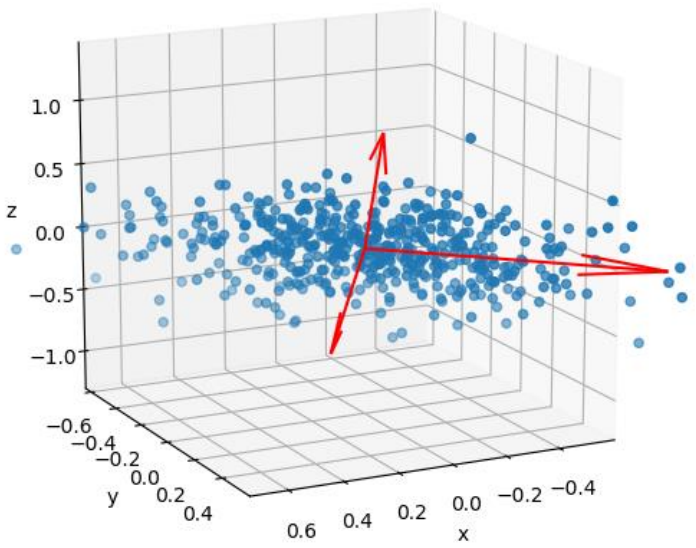
Workshop 10

COMP90051 Machine Learning

Semester 2, 2018

Refresher on PCA

- Given data matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$ (assume centred)
- Each principal component $\mathbf{p}_i$ should:
  - account for the greatest variance, i.e. maximise $\mathbf{p}_i^T \mathbf{X}^T \mathbf{X} \mathbf{p}_i$
  - be of unit length, i.e. $\mathbf{p}_i^T \mathbf{p}_i = 1$
  - be orthogonal to the previous components $\{\mathbf{p}_1, \ldots, \mathbf{p}_{i-1}\}$
- Solution: the $\mathbf{p}_i$'s are the eigenvectors of $\mathbf{X}^T \mathbf{X}$, sorted in decreasing order by eigenvalue



Learning Outcomes

At the end of this workshop you should be able to:

1. implement PCA using singular value decomposition (SVD)
2. apply PCA to visualise high-dimensional data in 3D
3. interpret PCA as a data compression algorithm

Worksheet 10

QoCT survey

- Your opportunity to provide anonymous feedback
- Teaching staff won't see the feedback until final results are released
- Please visit the link below (supports mobile devices)

http://go.unimelb.edu.au/4jx6