Data Science Toolbox Question Sheet

03.1 Latent Spaces and PCA

Daniel Lawson

Block 3

Short questions

1. If we use the identity that $C = Cov(X) = (1/(n-1))XX^T$, what assumptions have we made about X?
2. We project the data X onto the subspace U using the projection P using XP. What can you say about $(XP)P$?
3. In what sense is $\text{Cov}(X) = U\Sigma U^T$ truncated to low rank K the "best" low rank summary of X?
4. Give a high-level explanation for why minimising the mean squared error and maximising the variance of a low-dimensional representation of a matrix X into U, D and V leads to the same representation, the SVD.
5. What is an eigenvalue? What is an eigenvector?
6. What is the variance explained by the k-th eigenvector?
7. What is the relationship between singular values and eigenvalues?