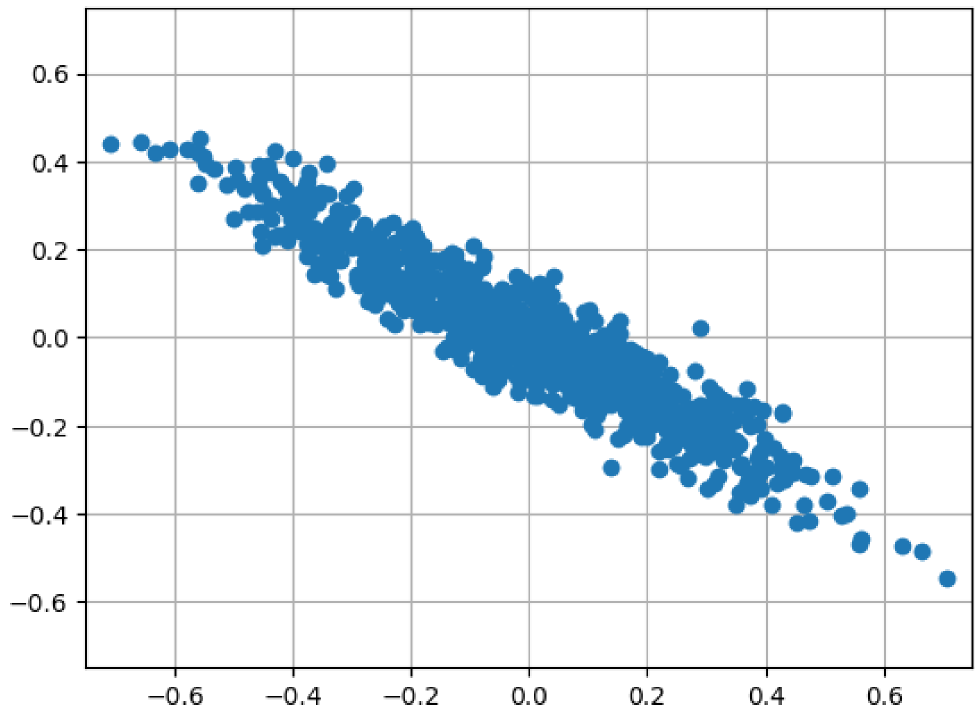
Linear Data Chapter 12

Written by: Emily J. King

1. Consider the zero-centered data cloud pictured below.



Let $\mathbf{X}$ be the 2×1000 zero-centered data matrix representing the data cloud. Which of the following are most likely to be the singular values of $\mathbf{X}$? You must justify your response with at least one complete sentence.

A. $\sigma_1 = 5$, $\sigma_2 = 4$

B. $\sigma_1 = 9$, $\sigma_2 = 1.5$

C. $\sigma_1 = 8$, $\sigma_2 = -1$

The points are very noticeably spread along one direction (if you're curious: approx slope -0.7265) than in the orthogonal direction. In [A.], the points would spread away *from the line almost as much as they spread along the line, which is not the case here. Singular values are non-negative, so [C.] cannot be correct. Thus, [B.], which has two fairly different singular values, must be correct.*

2. Let $\mathbf{U\Sigma V}^\top$ be the singular value decomposition of a $d \times n$ matrix $\mathbf{B}$, where $\mathbf{U}$ is a $d \times d$ orthogonal matrix with columns $\vec{u}_i$, $\mathbf{\Sigma}$ is diagonal $d \times n$ matrix with diagonal entries σ_i in non-increasing order, and $\mathbf{V}$ is a $n \times n$ orthogonal matrix with columns $\vec{v}_i$. Assume the singular values are all distinct. What is a correct formula for the best rank 1 approximation of $\mathbf{B}$? You must justify your response with at least one complete sentence.

   A. $\vec{u}_d \sigma_1 \vec{v}_n^\top$

   B. $\vec{u}_1 \sigma_1 \vec{v}_1^\top$

   C. $\mathbf{U\Sigma V}^\top$

   By Schmidt-Mirsky-Eckart-Young, the answer is [B]. Unless all of the singular values except one are zero, [C] will never be rank 1. And [A.] could only be correct if both $d = n$ and all of the singular values were equal (which we assumed wasn't true). In that case, there would be multiple best rank 1 approximations.

3. Name at least one application of PCA from the textbook or some other source. In the latter case, list the source.
   In the textbook: fitting affine subspaces to data, reducing dimensionality of geometric feature vectors of cloud images to allow for clustering, separating out artifacts from retinal images, generating low dimensional approximations of face images to allow for easy identification

4. Assume that $\mathbf{A}$ is a 3×10 matrix and that $\mathbf{AA}^\top$ has the eigenvalues $1, 4, 25$.

   (a) Explicitly list with justification the eigenvalues of $\mathbf{A}^\top\mathbf{A}$ (including multiplicity).
   We know that $\mathbf{A}^\top\mathbf{A}$ is 10×10 and has the same non-zero eigenvalues of $\mathbf{AA}^\top$. Thus, the eigenvalues of $\mathbf{A}^\top\mathbf{A}$ are

   $$0, 0, 0, 0, 0, 0, 0, 1, 4, 25$$

   (b) Explicitly list with justification the singular values of $\mathbf{A}$.
   Since $\mathbf{A}$ is 3×10, it has $3 = \min\{3, 10\}$ singular values which are the square roots of the eigenvalues of $\mathbf{AA}^\top$, i.e., $1, 2, 5$.