CAS CS 365
Lab 9

1. Eigenvalues and Eigenvectors
   a. Theorem 12.8: If matrices A and B are similar, i.e., if there is an invertible matrixP such that A = P^-1 * B * P, then they have the same eigenvalues

      Proof: For any eigenvalue and eigenvector pair (λ, x) , we know

      $Ax = \lambda x = P^{-1}BPx$, thus $\lambda Px = BPx$. Therefore (λ, Px)

   b. is a pair of eigenvalue and eigenvector of B.
   c. Definition: A matrix A is diagonalizable if A is similar to a diagonal matrix
   d. Theorem: A is diagonalizable if and only if A has n linearly independent eigenvectors

      Sketch:
      1. If A is diagonalizable, then there is an invertible matrix P and a diagonal matrix D, such that $D = P^{-1}AP$, thus PD=AP.
      2. Consider P as $[p_1, p_2, \ldots p_n]$, and D has elements $\lambda_1, \ldots \lambda_n$ on the diagonal, then $Ap_i = \lambda_i p_i$.
      3. P is invertible, so its columns are linearly independent.
      4. Assume A has n linearly independent eigenvectors, and reverse the steps above.
   e.
   f. Theorem 12.10. Let A be a real symmetric matrix, then all its eigenvalues and eigenvectors are real. Besides, A is orthogonally diagonalizable and

      $A = VDV^T = \sum_i \lambda_i v_i v_i^T$, where V is a matrix with eigenvectors as columns and D is a diagonal matrix with corresponding eigenvalues
2. SVD
   a. Singular value decomposition for matrix M

      $$M = U\Sigma V^T$$

   b. U and V are unitary matrices, meaning that rows/columns of U are orthonormal
   c. Sigma is a rectangular diagonal matrix with singular values on the diagonal
3. PCA
   a. Consider a n by p matrix X with column-wise zero empirical mean
      i. Each column can be considered as a feature
      ii. Each row is a data sample
   b. First round:
      i. We want to find a unit vector such that the projections of data samples on this vector have the largest variance

         $$w_{(1)} = \arg\max_{||w||=1} \left\{ \sum_{i=1}^{n} (x_i \cdot w)^2 \right\} = \arg\max_{||w||=1} \left\{ ||Xw||^2 \right\} = \arg\max_{||w||=1} \left\{ w^T X^T X w \right\}$$

c. K-th round

    i. We can form a new data matrix by subtracting all previous principal components from X

$$X_k = X - \sum_{s=1}^{k-1} X w_{(s)} w_{(s)}^T$$

    ii. Then repeat the process of finding the unit vector that leads to the max variance of projections

d. A matrix W can be formed as $[w_{(1)}|\dots|w_{(l)}]$, where l <= p. And the final transformed data is T = XW

e. Connection to SVD

    i. By SVD, we know that

$$X^T X = V\Sigma^T U^T U \Sigma V^T = V\Sigma^T \Sigma V^T$$
$$= V\hat{\Sigma}^2 V^T$$, where $\hat{\Sigma}^2$ is a diagonal matrix.

    ii. This is the format of eigen-decomposition, which implies the right singular vectors V of X are also the eigenvectors of X^T*X i.e., V= W, exactly the solution we need for PCA