Singular Value Decomposition X. Du

- Motivation: we know that we can decompose a symmetric matrix into $A = QDQ^T$ using eigenvalue and eigenvector analysis.
- As it turns out, we can also decompose any $m \times n$ matrix $A = U\Sigma V^T$.
  - $A = U\Sigma V^T$ is the **singular value decomposition**.
  - U is an $m \times m$ orthogonal matrix and V is an $n \times n$ orthogonal matrix.
  - Σ is an $m \times n$ block matrix of the form $\Sigma = \begin{pmatrix} D & O \\ O & O \end{pmatrix}$, where D is a diagonal matrix of **singular values** σ_i.
  - The column vectors of U are the **left-singular vectors** and the column vectors of V are the **right-singular vectors**.
  - For positive semi-definite matrices, the singular values and the singular vectors are the same as the eigenvalues and eigenvectors.
- For any $m \times n$ matrix A, consider A^TA.
  - A^TA is symmetric.
  - The singular values of A squared are the eigenvalues of A^TA. All eigenvalues of A^TA are real and non-negative, since A^TA is positive semi-definite.
  - Note the utility of A^TA. It can be used to determine whether or not a matrix has linearly independent column vectors.
- Finding the singular value decomposition:
  - Note that $A^TA = V\Sigma^TU^TU\Sigma V^T = V\Sigma^T\Sigma V^T$. $\Sigma^T\Sigma$ is just an $m \times m$ diagonal matrix with the diagonal entries as the singular values squared.
  - Note that $AV = U\Sigma$, so $A\vec{v}_i = \vec{u}_i\sigma_i$.
  - Step 1: Compute A^TA.
  - Step 2: Compute the singular values σ_i of A by taking the square roots of the eigenvalues of A^TA.
  - Step 3: Compute the right-singular vectors $\vec{v}_i$ in V by finding the normalized eigenvectors of A^TA.
  - Step 4: Compute the left-singular vectors $\vec{u}_i$ in U by using $\vec{u}_i = \frac{1}{\sigma_i}A\vec{v}_i$.
    - Alternatively, you can find the normalized eigenvectors of AA^T.
- Application: **Principal Component Analysis (PCA)**
  - Suppose you have vectors $\vec{x}_1, ..., \vec{x}_m \in \mathbb{R}^n$ that you want to compress into k-dimensions, where $k \leq n$. The set of data is also centered, i.e. for each $j \in [1, m]$, $E[\vec{x}_{ij}] = 0$ across all i.
  - Put these vectors as columns of matrix X. The mean of each row should be 0.
  - PCA aims to find the k directions in the data with the greatest variance, which would preserve the most information possible.
  - Find the SVD of X, i.e. find $X = U\Sigma V^T$. Note that $X = \sigma_1\vec{u}_1\vec{v}_1^T + \sigma_2\vec{u}_2\vec{v}_2^T + ...$, where $\vec{u}_i$ and $\vec{v}_i$ are the ith left and right-singular vectors respectively.
  - Take the k largest singular values with their associated singular vectors. These left-singular vectors are the principal components in PCA, and the rows of V represent the compressed data. To decompress the data, compute $X' = \sigma_1\vec{u}_1\vec{v}_1^T + ... + \sigma_k\vec{u}_k\vec{v}_k^T$.