Low Rank Approximations: Part I

BIOE 210
Review

We can uniquely decompose a vector $\mathbf{x}$ over a basis by finding the coefficients $\mathbf{a}$ such that

$$\mathbf{x} = a_1 \mathbf{v}_1 + \cdots + a_n \mathbf{v}_n$$

We find the coefficients by collecting the basis vectors into a matrix

$$\mathbf{V} = (\mathbf{v}_1 \mathbf{v}_2 \cdots \mathbf{v}_n)$$

and solving the system of equations

$$\mathbf{V} \mathbf{a} = \mathbf{x}$$

Using $\mathbf{V}$ and its inverse $\mathbf{V}^{-1}$ we can jump between the original vector and the coefficients of the decomposition.
The Singular Value Decomposition (SVD)

Any $m \times n$ matrix $A$ can be decomposed into the product of three matrices

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

- $U$ is an orthogonal $m \times m$ matrix.
- $\Sigma$ is a diagonal $m \times n$ matrix.
- $V$ is an orthogonal $n \times n$ matrix.
The Singular Value Decomposition (SVD)

Any $m \times n$ matrix $A$ can be decomposed into the product of three matrices

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

- $U$ is an orthogonal $m \times m$ matrix.
- $\Sigma$ is a diagonal $m \times n$ matrix.
- $V$ is an orthogonal $n \times n$ matrix.

Example: $2 \times 3$ matrix $A$:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{pmatrix} \begin{pmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{pmatrix}^T$$
What happens during matrix multiplication?

Let’s talk about multiplication using a $2 \times 3$ matrix $A$:

$$y = Ax$$

The input vector $x$ has three dimensions but the output vector $y$ has two dimensions.
What happens during matrix multiplication?

Let’s talk about multiplication using a $2 \times 3$ matrix $A$:

$$y = Ax$$

The input vector $x$ has three dimensions but the output vector $y$ has two dimensions.

- What happens to the third dimension?
- Is that information lost?
- If not, how does $A$ compress the information from $x$ into the smaller vector $y$?
- How many times are we going to talk about matrix multiplication?
Every matrix has an input and output basis

\[
\begin{pmatrix}
  a_{11} & a_{12} & a_{13} \\
  a_{21} & a_{22} & a_{23}
\end{pmatrix} =
\begin{pmatrix}
  u_{11} & u_{12} \\
  u_{21} & u_{22}
\end{pmatrix}
\begin{pmatrix}
  \sigma_1 & 0 & 0 \\
  0 & \sigma_2 & 0
\end{pmatrix}
\begin{pmatrix}
  v_{11} & v_{12} & v_{13} \\
  v_{21} & v_{22} & v_{23} \\
  v_{31} & v_{32} & v_{33}
\end{pmatrix}^{T}
\]

- The columns of \( \mathbf{V} \) are an orthonormal basis for the \textit{input space} of \( \mathbf{A} \).
- The columns of \( \mathbf{U} \) are an orthonormal basis for the \textit{output space} of \( \mathbf{A} \).
- Matrix multiplication is simply a change of basis from the input to output spaces. The switch happens in the matrix \( \Sigma \).
Step 1: Decompose $\mathbf{x}$ onto the input basis $\mathbf{V}$

- We want to find coefficients $\mathbf{a}$ such that $\mathbf{x}$ can be expressed using the input basis (the columns of $\mathbf{V}$).
- We can find these coefficients using $\mathbf{V}^{-1}\mathbf{x}$.
- But, $\mathbf{V}$ is an orthogonal matrix, so $\mathbf{V}^{-1} = \mathbf{V}^T$. The coefficients to decompose $\mathbf{x}$ onto $\mathbf{V}$ are simply $\mathbf{V}^T\mathbf{x}$. 
Step 1: Decompose \( \mathbf{x} \) onto the input basis \( \mathbf{V} \)

- We want to find coefficients \( \mathbf{a} \) such that \( \mathbf{x} \) can be expressed using the input basis (the columns of \( \mathbf{V} \)).
- We can find these coefficients using \( \mathbf{V}^{-1}\mathbf{x} \).
- But, \( \mathbf{V} \) is an orthogonal matrix, so \( \mathbf{V}^{-1} = \mathbf{V}^T \). The coefficients to decompose \( \mathbf{x} \) onto \( \mathbf{V} \) are simply \( \mathbf{V}^T\mathbf{x} \).

\[
\begin{align*}
\mathbf{y} &= \mathbf{Ax} \\
&= \mathbf{U}\Sigma\mathbf{V}^T\mathbf{x} \\
&= \mathbf{U}\Sigma\mathbf{a}
\end{align*}
\]

\( \mathbf{a} = \mathbf{V}^T\mathbf{x} \) are the coefficients that decompose \( \mathbf{x} \) onto the input basis.
Step 2: Rescale the input coefficients to match the output basis

\[
\begin{pmatrix}
  a_{11} & a_{12} & a_{13} \\
  a_{21} & a_{22} & a_{23}
\end{pmatrix}
\begin{pmatrix}
  x_1 \\
  x_2 \\
  x_3
\end{pmatrix} =
\begin{pmatrix}
  u_{11} & u_{12} \\
  u_{21} & u_{22}
\end{pmatrix}
\begin{pmatrix}
  \sigma_1 & 0 & 0 \\
  0 & \sigma_2 & 0
\end{pmatrix}
\begin{pmatrix}
  v_{11} & v_{12} & v_{13} \\
  v_{21} & v_{22} & v_{23} \\
  v_{31} & v_{32} & v_{33}
\end{pmatrix}^T
\begin{pmatrix}
  x_1 \\
  x_2 \\
  x_3
\end{pmatrix}
\]

\[
\begin{pmatrix}
  y_1 \\
  y_2
\end{pmatrix} =
\begin{pmatrix}
  u_{11} & u_{12} \\
  u_{21} & u_{22}
\end{pmatrix}
\begin{pmatrix}
  \sigma_1 & 0 & 0 \\
  0 & \sigma_2 & 0
\end{pmatrix}
\begin{pmatrix}
  a_1 \\
  a_2 \\
  a_3
\end{pmatrix}
\]

\[
= \begin{pmatrix}
  u_{11} & u_{12} \\
  u_{21} & u_{22}
\end{pmatrix}
\begin{pmatrix}
  \sigma_1 a_1 \\
  \sigma_2 a_2
\end{pmatrix}
\]
Step 3: Reconstruct $y$ using the output basis

\[
\begin{pmatrix}
y_1 \\
y_2
\end{pmatrix} =
\begin{bmatrix}
u_{11} & u_{12} \\
u_{21} & u_{22}
\end{bmatrix}
\begin{pmatrix}
\sigma_1 a_1 \\
\sigma_2 a_2
\end{pmatrix}
\]

\[
y = \sigma_1 a_1 u + \sigma_2 a_2 u_2
\]
The information hierarchy in the SVD

The singular values ($\sigma_i$) map the right singular vectors ($v_i$) to the left singular vectors ($u_i$). This mapping happens in the matrix $\Sigma$. Any extra right or left singular vectors are “zeroed-out" by $\Sigma$.

$$u_1 \leftarrow \sigma_1 v_1$$

$$u_m \leftarrow \sigma_m v_m$$

$$0 \leftarrow 0 v_{m+1}$$

$$0 \leftarrow 0 v_n$$

All singular vectors are unit vectors, so the largest singular values identify the most important parts of the matrix.
The information hierarchy in the SVD

The singular values ($\sigma_i$) map the right singular vectors ($v_i$) to the left singular vectors ($u_i$). This mapping happens in the matrix $\Sigma$. Any extra right or left singular vectors are "zeroed-out" by $\Sigma$.

\[
\begin{align*}
    u_1 & \leftarrow \sigma_1 \quad v_1 \\
    \vdots \\
    u_m & \leftarrow \sigma_m \quad v_m \\
    0 & \leftarrow 0 \quad v_{m+1} \\
    \vdots \\
    0 & \leftarrow 0 \quad v_n
\end{align*}
\]

All singular vectors are unit vectors, so the largest singular values identify the most important parts of the matrix.