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

$$\mathbf{V} = \begin{pmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\end{pmatrix}$$

and solving the system of equations

$$\mathbf{Va} = \mathbf{x}$$

Using $\mathbf{V}$ and its inverse $\mathbf{V}^{-1}$ we can jump between the original vector and the coefficients of the decomposition.

$$\begin{tikzcd} \mathbf{x} \arrow[r, bend left, "\mathbf{V}^{-1}"] & \mathbf{a} \arrow[l, bend left, "\mathbf{V}"] \end{tikzcd}$$

The Singular Value Decomposition (SVD)

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

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathsf{T}}$$

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

The Singular Value Decomposition (SVD)

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

$$\mathbf{A} = \mathbf{U\Sigma V}^\mathsf{T}$$

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

Example: 2×3 matrix $\mathbf{A}$:

$$\underbrace{\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}}_{\mathbf{A}} = \underbrace{\begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}}_{\mathbf{U}} \underbrace{\begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{pmatrix}}_{\mathbf{\Sigma}} \underbrace{\begin{pmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{pmatrix}^\mathsf{T}}_{\mathbf{V}^\mathsf{T}}$$

What happens during matrix multiplication?

Let's talk about multiplication using a 2×3 matrix $\mathbf{A}$:

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

The input vector $\mathbf{x}$ has three dimensions but the output vector $\mathbf{y}$ has two dimensions.

What happens during matrix multiplication?

Let's talk about multiplication using a 2×3 matrix $\mathbf{A}$:

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

The input vector $\mathbf{x}$ has three dimensions but the output vector $\mathbf{y}$ has two dimensions.

- What happens to the third dimension?
- Is that information lost?
- If not, how does $\mathbf{A}$ compress the information from $\mathbf{x}$ into the smaller vector $\mathbf{y}$?
- How many times are we going to talk about matrix multiplication?

Every matrix has an input and output basis

$$\underbrace{\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}}_{\mathbf{A}} = \underbrace{\begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}}_{\mathbf{U}} \underbrace{\begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{pmatrix}}_{\mathbf{\Sigma}} \underbrace{\begin{pmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{pmatrix}^{\mathsf{T}}}_{\mathbf{V}^{\mathsf{T}}}$$

- The columns of **V** are an orthonormal basis for the *input space* of **A**.
- The columns of **U** are an orthonormal basis for the *output space* of **A**.
- Matrix multiplication is simply a change of basis from the input to output spaces. The switch happens in the matrix $\mathbf{\Sigma}$.

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 $\mathbf{V}^{-1} = \mathbf{V}^{\mathsf{T}}$. The coefficients to decompose $\mathbf{x}$ onto $\mathbf{V}$ are simply $\mathbf{V}^{\mathsf{T}}\mathbf{x}$.

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 $\mathbf{V}^{-1} = \mathbf{V}^{\mathsf{T}}$. The coefficients to decompose $\mathbf{x}$ onto $\mathbf{V}$ are simply $\mathbf{V}^{\mathsf{T}}\mathbf{x}$.

$$\begin{aligned}\mathbf{y} &= \mathbf{A}\mathbf{x} \\ &= \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{T}}\mathbf{x} \\ &= \mathbf{U}\boldsymbol{\Sigma}\mathbf{a}\end{aligned}$$

$\mathbf{a} = \mathbf{V}^{\mathsf{T}}\mathbf{x}$ are the coefficients that decompose $\mathbf{x}$ onto the input basis.

Step 2: Rescale the input coefficients to match the output basis

$$\underbrace{\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}_{\mathbf{x}} = \underbrace{\begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}}_{\mathbf{U}} \underbrace{\begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{pmatrix}}_{\mathbf{\Sigma}} \underbrace{\begin{pmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{pmatrix}^{\mathsf{T}}}_{\mathbf{V}^{\mathsf{T}}} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}_{\mathbf{x}}$$

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}}_{\mathbf{U}} \underbrace{\begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{pmatrix}}_{\mathbf{\Sigma}} \underbrace{\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}}_{\mathbf{a}}$$

$$= \underbrace{\begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}}_{\mathbf{U}} \underbrace{\begin{pmatrix} \sigma_1 a_1 \\ \sigma_2 a_2 \end{pmatrix}}_{\mathbf{\Sigma a}}$$

Step 3: Reconstruct $\mathbf{y}$ using the output basis

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}}_{\mathbf{U}} \underbrace{\begin{pmatrix} \sigma_1 a_1 \\ \sigma_2 a_2 \end{pmatrix}}_{\mathbf{\Sigma a}}$$

$$= \underbrace{\begin{pmatrix} \sigma_1 a_1 u_{11} + \sigma_2 a_2 u_{12} \\ \sigma_1 a_1 u_{21} + \sigma_2 a_2 u_{22} \end{pmatrix}}_{\mathbf{U\Sigma a}}$$

$$\mathbf{y} = \sigma_1 a_1 \mathbf{u}_1 + \sigma_2 a_2 \mathbf{u}_2$$

The information hierarchy in the SVD

The singular values (σ_i) map the right singular vectors ($\mathbf{v}_i$) to the left singular vectors ($\mathbf{u}_i$). This mapping happens in the matrix $\mathbf{\Sigma}$. Any extra right or left singular vectors are “zeroed-out" by $\mathbf{\Sigma}$.

$$\mathbf{u}_1 \xleftarrow{\sigma_1} \mathbf{v}_1$$

$$\vdots$$

$$\mathbf{u}_m \xleftarrow{\sigma_m} \mathbf{v}_m$$

$$\mathbf{0} \xleftarrow{0} \mathbf{v}_{m+1}$$

$$\vdots$$

$$\mathbf{0} \xleftarrow{0} \mathbf{v}_n$$

The information hierarchy in the SVD

The singular values (σ_i) map the right singular vectors ($\mathbf{v}_i$) to the left singular vectors ($\mathbf{u}_i$). This mapping happens in the matrix $\mathbf{\Sigma}$. Any extra right or left singular vectors are “zeroed-out" by $\mathbf{\Sigma}$.

$$
\begin{array}{ccc}
\mathbf{u}_1 & \xleftarrow{\sigma_1} & \mathbf{v}_1 \\
 & \vdots & \\
\mathbf{u}_m & \xleftarrow{\sigma_m} & \mathbf{v}_m \\
\mathbf{0} & \xleftarrow{0} & \mathbf{v}_{m+1} \\
 & \vdots & \\
\mathbf{0} & \xleftarrow{0} & \mathbf{v}_n
\end{array}
$$

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