SVD分解的真正理解：几何角度

旋转矩阵(LLM旋转位置编码的前提)

教学演示：用复数推导逆时针旋转矩阵

1. 建立对应关系

首先，我们把二维平面上的一个点（向量）$\begin{bmatrix} x \\ y \end{bmatrix}$ 对应到一个复数 z：

$z = x + iy$

- 实部 x 对应横坐标。
- 虚部 y 对应纵坐标。

2. 旋转的神器：欧拉公式

在复数世界里，**旋转**就是一个乘法操作。
如果要把复数 z 逆时针旋转 θ 度，只需要乘以一个模长为 1 的复数 $e^{i\theta}$。

根据欧拉公式：
$e^{i\theta} = \cos\theta + i\sin\theta$

所以，新的位置 z' 等于旧的位置 z 乘以旋转因子：
$z' = z \cdot (\cos\theta + i\sin\theta)$

3. 展开计算（代数推导）

现在我们把 $z = x + iy$ 代进去，看看结果是什么。
设旋转后的新坐标为 (x', y')，即 $z' = x' + iy'$。

$x' + iy' = (x + iy)(\cos\theta + i\sin\theta)$

我们要做的就是简单的多项式乘法（注意 $i^2 = -1$）：

$$\begin{aligned} x' + iy' &= x\cos\theta + ix\sin\theta + iy\cos\theta + i^2 y\sin\theta \\ &= x\cos\theta + ix\sin\theta + iy\cos\theta - y\sin\theta \end{aligned}$$

现在，我们将**实部**（不带 i 的）和**虚部**（带 i 的）分开归类：

- **实部 (Real Part):** $x' = x\cos\theta - y\sin\theta$
- **虚部 (Imaginary Part):** $y' = x\sin\theta + y\cos\theta$

4. 变回矩阵形式

看上面的实部和虚部方程组：

$$\begin{cases} x' = (\cos\theta)\cdot x + (-\sin\theta)\cdot y \\ y' = (\sin\theta)\cdot x + (\cos\theta)\cdot y \end{cases}$$

在线性代数中，这正是矩阵乘法的定义。我们将系数提取出来，写成矩阵形式：

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

证毕!

中间那个矩阵：
$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$
正是标准的**逆时针旋转矩阵。**

第一部分：基础工具箱（公式速查）

在二维平面中，线性变换主要靠这两种基本矩阵：

1. 旋转矩阵 (Rotation Matrix)

如果你想把一个点 (x, y) 绕原点旋转 θ 度，你需要乘以旋转矩阵 R。

- **逆时针旋转 (Counter-Clockwise, CCW) θ 度:**
  $R_{ccw} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

2. 缩放矩阵 (Scaling Matrix)

如果你想把 x 坐标拉伸 s_x 倍，把 y 坐标拉伸 s_y 倍，你需要乘以对角矩阵 S。

$$S = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix}$$

从圆到椭圆——看见矩阵背后的SVD

1. 核心直觉

在线性代数中，一个 2×2 的矩阵 A 就像是一个"变形金刚"。
如果你在纸上画一个**单位圆**（半径为1的圆），然后用矩阵 A 去乘以圆上的每一个点，这个圆通常会变成一个**椭圆**。

SVD 分解公式 $A = U\Sigma V^T$ 告诉了我们这个变形动作是分三步完成的：

1. V^T **(旋转/反射)**：先把圆旋转一下，让圆上某些特殊的点对准坐标轴。
2. Σ **(缩放)**：沿着坐标轴方向拉伸或压缩。**这一步把圆变成了椭圆。**
3. U **(旋转/反射)**：把这个刚刚生成的椭圆再旋转到最终的位置。

一句话总结：SVD 告诉了我们这个椭圆的长轴和短轴有多长（Σ），以及它们指向哪里（U）。

---

2. 动手算一算：一个简单的数值例子

为了看清过程，我们不直接给一个乱七八糟的矩阵 A，而是我们**自己构造**一个 A，看看它是如何把圆变成椭圆的。

假设我们想制造一个这样的变换：

- **第一步**：顺时针旋转 45°。
- **第二步**：横向拉长 3 倍，纵向拉长 1 倍。
- **第三步**：逆时针旋转 90°。

根据 SVD 公式 $A = U\Sigma V^T$，我们来定义这三个矩阵。

准备工作（为了计算方便，取 $\frac{1}{\sqrt{2}} \approx 0.7$）

1. **定义 V^T (初次旋转)**：
   为了方便理解，我们选取一个特殊的向量作为追踪对象。
   假设单位圆上有一个点 $x = \begin{bmatrix} 0.7 \\ 0.7 \end{bmatrix}$（位于 45° 方向）。
   我们定义 V^T 为顺时针旋转 45° 的矩阵：
   $V^T = \begin{bmatrix} 0.7 & 0.7 \\ -0.7 & 0.7 \end{bmatrix}$
   (注：这是旋转矩阵，正交矩阵)
2. **定义 Σ (缩放)**：
   奇异值矩阵是对角矩阵。这里 $\sigma_1 = 3$ (长轴)，$\sigma_2 = 1$ (短轴)。
   $\Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$
3. **定义 U (最终旋转)**：
   我们想把拉伸后的图形逆时针旋转 90°。
   $U = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$

---

3. 见证奇迹：追踪一个点的旅程

现在，我们要计算 Ax，也就是 $U\Sigma V^T x$。我们将目光死死盯着圆上的那个点 $x = \begin{bmatrix} 0.7 \\ 0.7 \end{bmatrix}$。

Step 1: V^T 进场（调整姿态）
$$V^T x = \begin{bmatrix} 0.7 & 0.7 \\ -0.7 & 0.7 \end{bmatrix} \begin{bmatrix} 0.7 \\ 0.7 \end{bmatrix} = \begin{bmatrix} 0.49 + 0.49 \\ -0.49 + 0.49 \end{bmatrix} \approx \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

- **发生了什么？** 原始向量 x 指向 45°。V^T 把它旋转到了 X 轴上 $(1, 0)$。
- **意义：** V 的列向量（即 V^T 的行）代表了**"圆上哪些方向会被拉伸得最长或最短"**。在这里，它把我们将要拉伸的方向对准了 X 轴。

Step 2: Σ 进场（变形！）
$$\Sigma(V^T x) = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

- **发生了什么？** 单位圆变成了椭圆。原本长度为 1 的向量，现在变成了长度为 3。
- **意义：** 这里的 3 就是最大的奇异值 σ_1。它决定了椭圆**长半轴的长度。**

Step 3: U 进场（摆放位置）
$$U(\Sigma V^T x) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

- **发生了什么?** 这个被拉长的椭圆（目前躺在X轴上），被整体旋转了 $90°$，竖了起来。
- **意义：** 最终的向量长度依然是 3，但方向变了。U 的列向量决定了椭圆**最终主轴的方向**。

---

4. 总结：SVD 与 椭圆的一一对应

如果有一个矩阵 A，它的 SVD 分解是 $A = U\Sigma V^T$，那么 A 作用在单位圆上形成的椭圆具有以下特征：

1. **椭圆的长短轴长度**：完全由 Σ 对角线上的数字（奇异值 σ_1, σ_2）决定。
   - 例子中：长轴半径是 3，短轴半径是 1。
2. **椭圆长短轴的方向**：完全由矩阵 U 的列向量（u_1, u_2）决定。
   - 例子中：最终的长轴沿着 Y 轴方向（因为 $u_1 = [0, 1]^T$），短轴沿着 X 轴方向。
3. **谁变成了长轴**：圆上原本位于 V 的列向量（v_1）方向的点，被变换成了椭圆的长轴端点。

SVD分解的证明

我们来用纯粹的代数逻辑，不带任何具体数字，来一步步推导出SVD分解的存在性。为了便于理解，我们假设 A 是一个 3x3 的实数矩阵。

我们的目标：证明任何 3x3 实矩阵 A 都可以被分解为 $A = U\Sigma V^T$ 的形式，其中：

- **U** 是一个 3x3 的**正交矩阵** ($U^TU = I$)，它的列向量构成了输出空间的一组标准正交基。
- **V** 是一个 3x3 的**正交矩阵** ($V^TV = I$)，它的列向量构成了输入空间的一组标准正交基。
- **Σ** 是一个 3x3 的**对角矩阵**，对角线上的元素 $\sigma_1, \sigma_2, \sigma_3$ 是非负的，并按降序排列，称为**奇异值**。

第一部分：SVD分解的整体流程（算法视角）

在不知道为什么能这么做之前，我们先描述“怎么做”。这个流程本身就蕴含了证明的思路。

1. **构造对称矩阵**：计算 $M = A^TA$。这个矩阵 M 是一个 3x3 的实对称矩阵。
2. **寻找输入空间的基 (V)**：
   - 求解 M 的特征值 $\lambda_1, \lambda_2, \lambda_3$ 和对应的特征向量 v_1, v_2, v_3。
   - 因为 M 是实对称矩阵，我们总能找到一组相互**正交**的特征向量。
   - 将这些特征向量单位化，得到一组**标准正交基** $\{v_1, v_2, v_3\}$。
   - 用这些向量作为列，构建正交矩阵 $V = [v_1 \quad v_2 \quad v_3]$。
3. **确定奇异值 (Σ)**：
   - 我们稍后会证明 M 的所有特征值 λ_i 都是非负的。
   - 定义奇异值 $\sigma_i = \sqrt{\lambda_i}$。
   - 将它们按从大到小排列，并放在对角矩阵 Σ 的对角线上。
4. **寻找输出空间的基 (U)**：
   - 对于所有 $\sigma_i > 0$ 的情况，我们定义 $u_i = \frac{1}{\sigma_i}Av_i$。
   - 对于 $\sigma_i = 0$ 的情况，对应的 u_i 需要通过其他方法补全，以形成一组完整的标准正交基。
   - 用得到的标准正交基 $\{u_1, u_2, u_3\}$ 作为列，构建正交矩阵 $U = [u_1 \quad u_2 \quad u_3]$。
5. **组合**：最终，我们断言 $A = U\Sigma V^T$。

第二部分：从实对称矩阵的性质出发进行证明

现在，我们来证明为什么上述流程是可行且正确的。整个证明的核心基石是**实对称矩阵的谱定理 (Spectral Theorem)**。

谱定理核心内容：对于任何一个 n×n 的实对称矩阵 M，都存在一个由其特征向量构成的**标准正交基**，并且其所有特征值都是实数。

步骤 1：分析 A^TA 的性质

我们构造的矩阵是 $M = A^TA$。

- **对称性**： $M^T = (A^TA)^T = A^T(A^T)^T = A^TA = M$。所以 A^TA 是一个实对称矩阵。
- **特征值的非负性**：这是至关重要的一步。
   - 设 λ 是 A^TA 的一个特征值，对应的特征向量为 v（$v \neq \mathbf{0}$）。
   - 根据定义，我们有 $(A^TA)v = \lambda v$。
   - 我们用 v^T 左乘这个等式：$v^T(A^TA)v = v^T(\lambda v)$。

- 左边可以重新组合：$(Av)^T(Av) = ||Av||^2$。
- 右边可以提出标量 λ：$\lambda(v^Tv) = \lambda||v||^2$。
- 所以我们得到：$||Av||^2 = \lambda||v||^2$。
- 因为向量的范数平方 $||Av||^2$ 和 $||v||^2$ 都是非负的，且 $v \neq \mathbf{0}$ 意味着 $||v||^2 > 0$，所以我们必须有 $\lambda \geq 0$。

结论：A^TA 是一个实对称矩阵，且其所有特征值都是非负实数。这就保证了我们可以对特征值开平方得到实数奇异值 $\sigma_i = \sqrt{\lambda_i}$。

步骤 2：构造 V 和 Σ

根据谱定理，因为 A^TA 是实对称的，我们一定可以找到一组**标准正交**的特征向量 $\{v_1, v_2, v_3\}$，它们构成 $\mathbb{R}^3$ 的一组基。我们将它们作为列，构成正交矩阵 V。
同时，我们用对应的非负特征值的平方根 $\sigma_i = \sqrt{\lambda_i}$ 构成对角矩阵 Σ。

步骤 3：构造 U 并证明其正交性

这是证明中最巧妙的部分。我们定义了一组新的向量 $\{u_i\}$，现在需要证明它们也是标准正交的。

- 我们定义 $u_i = \frac{1}{\sigma_i}Av_i$ （暂时只考虑 $\sigma_i > 0$ 的情况）。
- 现在我们来计算任意两个这样的向量 u_i 和 u_j 的点积 $u_i^Tu_j$：
  $u_i^Tu_j = (\frac{1}{\sigma_i}Av_i)^T(\frac{1}{\sigma_j}Av_j) = \frac{1}{\sigma_i\sigma_j}(v_i^TA^T)(Av_j)$
  $u_i^Tu_j = \frac{1}{\sigma_i\sigma_j}v_i^T(A^TAv_j)$
- 我们知道 v_j 是 A^TA 的特征向量，所以 $A^TAv_j = \lambda_jv_j = \sigma_j^2v_j$。代入上式：
  $u_i^Tu_j = \frac{1}{\sigma_i\sigma_j}v_i^T(\sigma_j^2v_j) = \frac{\sigma_j^2}{\sigma_i\sigma_j}(v_i^Tv_j) = \frac{\sigma_j}{\sigma_i}(v_i^Tv_j)$
- 现在分情况讨论：
  - **如果** $i \neq j$: 因为 $\{v_i\}$ 是一组标准正交基，所以 $v_i^Tv_j = 0$。因此，$u_i^Tu_j = 0$。这意味着这些 u 向量是相互**正交**的。
  - **如果** $i = j$: 因为 $\{v_i\}$ 是单位向量，所以 $v_i^Tv_i = 1$。因此，$u_i^Tu_i = \frac{\sigma_i}{\sigma_i}(1) = 1$。这意味着每个 u 向量都是**单位向量**。

结论：我们通过 Av_i 构造出来的这组向量 $\{u_i\}$ (对应非零奇异值的部分) 自动就是**标准正交**的！

- **处理零奇异值**：如果存在某个 $\sigma_k = 0$，这意味着 $A^TAv_k = \mathbf{0}$，也意味着 $||Av_k||^2 = 0$，即 $Av_k = \mathbf{0}$。我们已经有了一组标准正交向量 $\{u_i\}$，如果它们不足以张成整个 $\mathbb{R}^3$ 空间，我们可以通过 Gram-Schmidt 等方法，添加新的标准正交向量来补全这组基，形成完整的 $\{u_1, u_2, u_3\}$。然后用它们构建正交矩阵 U。

步骤 4：整合与验证

我们已经建立了关系：

- $Av_i = \sigma_iu_i$ (对于 $\sigma_i > 0$)
- $Av_i = \mathbf{0}$ (对于 $\sigma_i = 0$)

我们可以将这两种情况统一写成矩阵形式。考虑整个矩阵 AV：
$AV = A\begin{bmatrix}v_1 & v_2 & v_3\end{bmatrix} = \begin{bmatrix}Av_1 & Av_2 & Av_3\end{bmatrix}$
$AV = \begin{bmatrix}\sigma_1u_1 & \sigma_2u_2 & \dots & \sigma_nu_n\end{bmatrix}$ (*这里 σ_iu_i 对于 $\sigma_i = 0$ 的情况就是零向量*)

现在我们再看 $U\Sigma$：

$$U\Sigma = \begin{bmatrix}u_1 & u_2 & u_3\end{bmatrix}\begin{pmatrix}\sigma_1 & 0 & 0\\ 0 & \sigma_2 & 0\\ 0 & 0 & \sigma_3\end{pmatrix}$$

$U\Sigma = \begin{bmatrix}\sigma_1u_1 & \sigma_2u_2 & \sigma_3u_3\end{bmatrix}$

比较两边，我们发现 $AV = U\Sigma$。

因为 V 是一个正交矩阵，所以 $V^T = V^{-1}$。我们在等式 $AV = U\Sigma$ 两边同时右乘 V^T：
$(AV)V^T = (U\Sigma)V^T$
$A(VV^T) = U\Sigma V^T$
$AI = U\Sigma V^T$
$A = U\Sigma V^T$

证明完毕。我们从“A^TA 是一个实对称矩阵”这一核心事实出发，利用谱定理保证了标准正交基 V 的存在，并推导出了另一组标准正交基 U 的存在，最终构成了SVD分解。