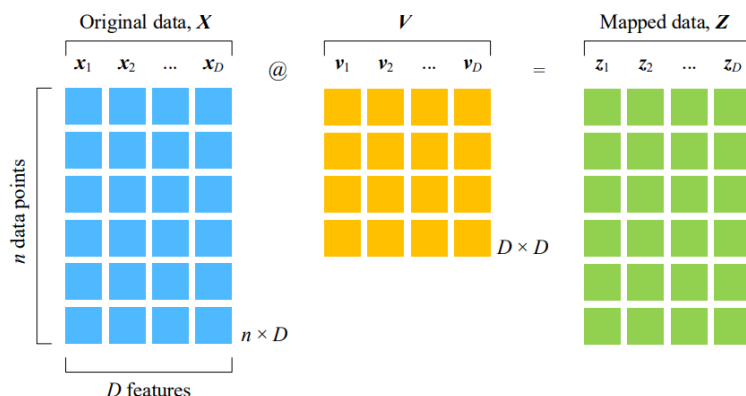


数据的正交投影

首先，我们考虑一个有趣的矩阵乘法运算——一个数据矩阵 X ，一个正交矩阵 V （这意味着它的列全部是规范正交基），那么结果中的每一个元素都是一行数据向着某一个正交基的投影长度：

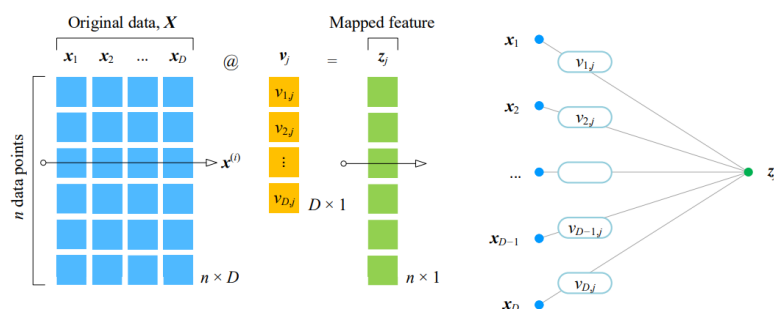


例如， $z_{1,1}$ 就是第一行数据向着第一列代表的标准正交基的投影长度（标量投影）。

这个正交投影可以从两个不同的角度来说明：

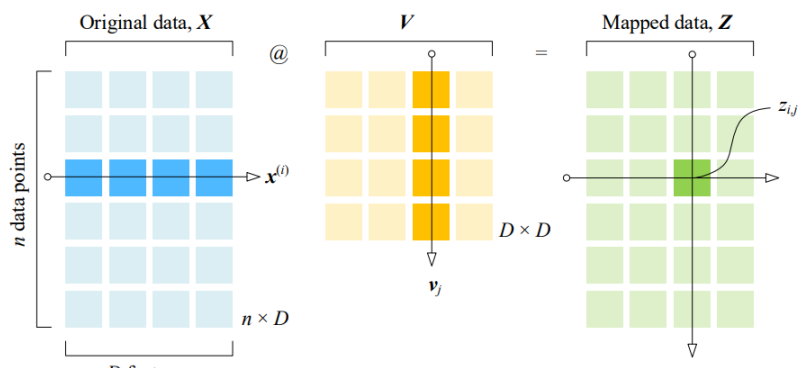
- 将 Z 和 V 写成列向量，此时，提取出 Z 中的一列，就是 $z_j = Xv_j$ ，这意味着 z_j 是 X 的各个列线性组合得到的。

注意：这个结论必须记住：一个列向量左乘一个矩阵得到一个向量，则得到的向量实际上是由矩阵的列向量线性组合而来的！

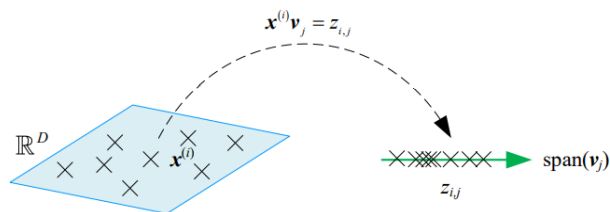


- 将 Z 和 X 写成行向量，此时，提取出 Z 中的一行，就是 $z_i = x_i V$ ，这意味着 Z 的一行是 X 的一行向着四个规范正交基上面投影得到的。

如图，展示了一行数据的投影过程：



这实际上是将 D 维上的数据打到了 1 维。



数据的“二次投影”：先利用投影展开，后叠加

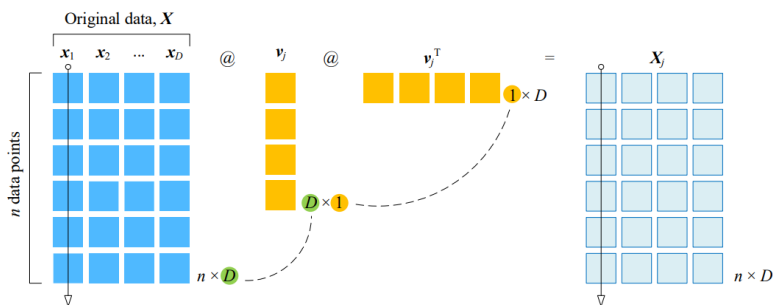
我们再考虑一个特殊的矩阵乘法：

$$X = XI = XVV^T$$

将 V 写成列向量的形式：

$$X = XVV^T = X[v_1 \ v_2 \ v_3 \ \cdots \ v_D][v_1^T \ \cdots \ v_D^T]^T = \sum_{i=1}^D Xv_i v_i^T$$

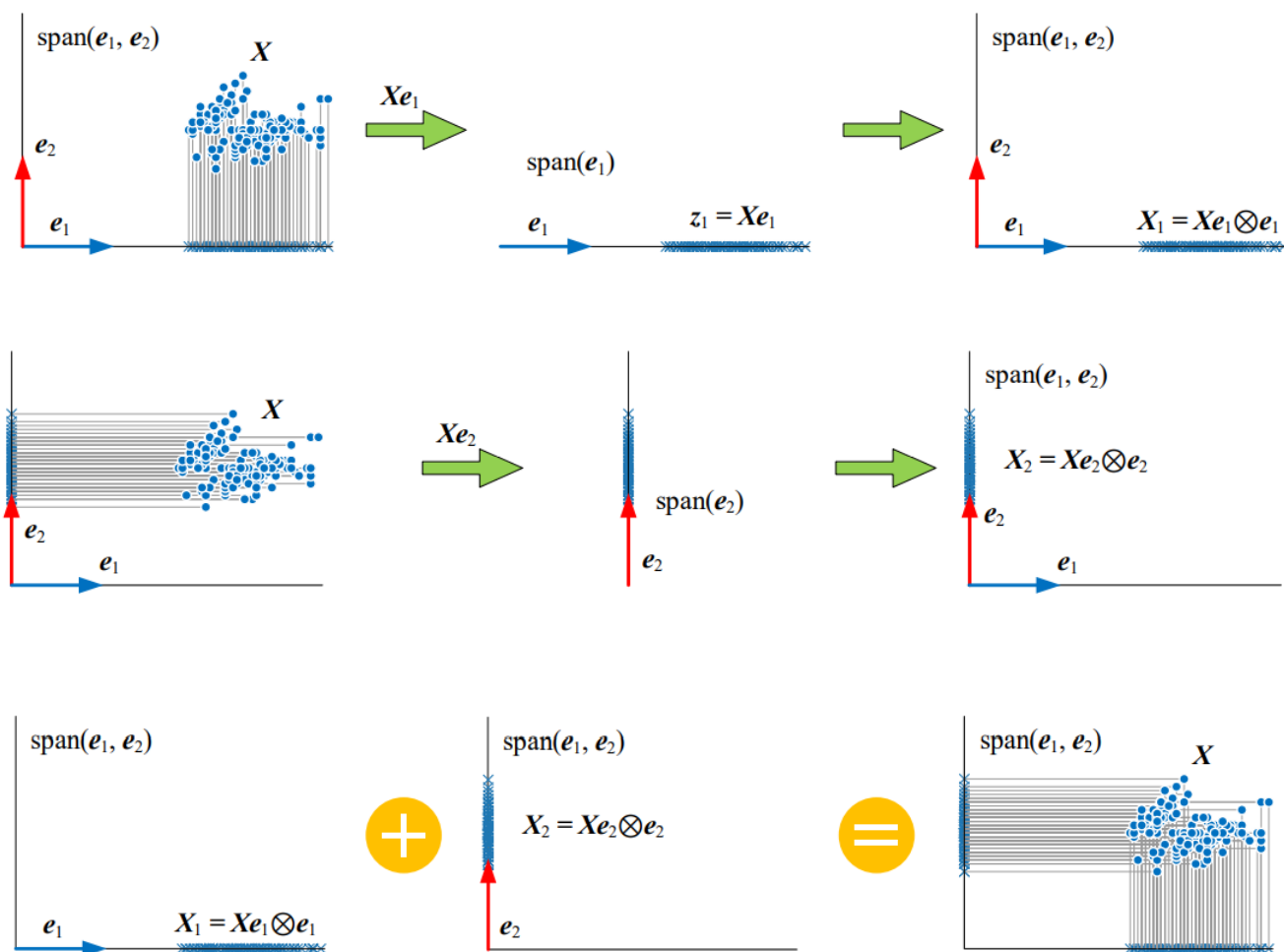
数据在这个过程中历经了如下的变化：首先， X 的列向量被线性组合，成为 $D \times 1$ 的列向量，此后，又被展开成与初始形状相同的矩阵。此后，多个矩阵层层叠加，又回到了初始的数据。



记 $X = X_1 + X_2 + \cdots X_D$ ，仅仅考虑 X_j 的一个行向量经历了何种运算：

$$x_j^{(i)} = x^{(i)} v_j v_j^T = z_{i,j} v_j^T$$

那么，这代表着 \mathbb{R}^D 中的数据点先投影到一维的空间 $\text{span}(v_j)$ 中，再通过某种变换回到 \mathbb{R}^D ，如果我们将再写一下，就发现 $X_j = Xv_j \otimes v_j$ ， X_j 实际上是数据矩阵 X 在 v_j 上的向量投影组成的矩阵，每一行都是一个投影向量。这个操作是要降维的，就算你把它形式还原了，这个降维也回不去。因此，每一个 X_j 都仅仅在一个方向上。



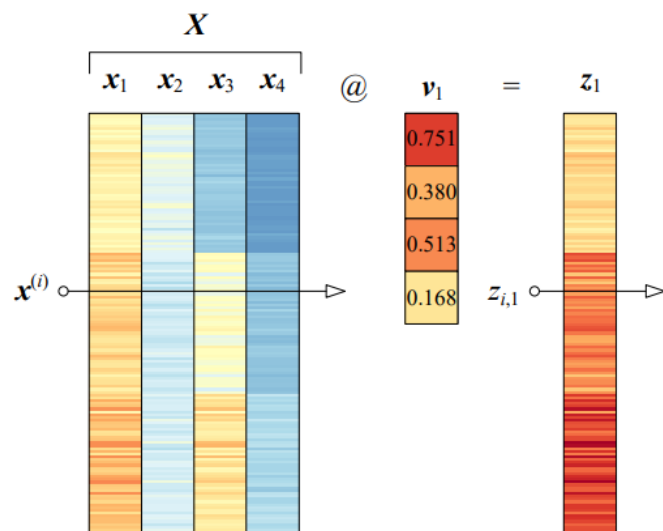
万里挑一的正交基

显然，我们的数据可以向着任意的一组正交基进行向量投影，之后再叠加起来。如果我们给出一组特殊的正交基呢？对于鸢尾花数据集，我们就给出下面这组基：

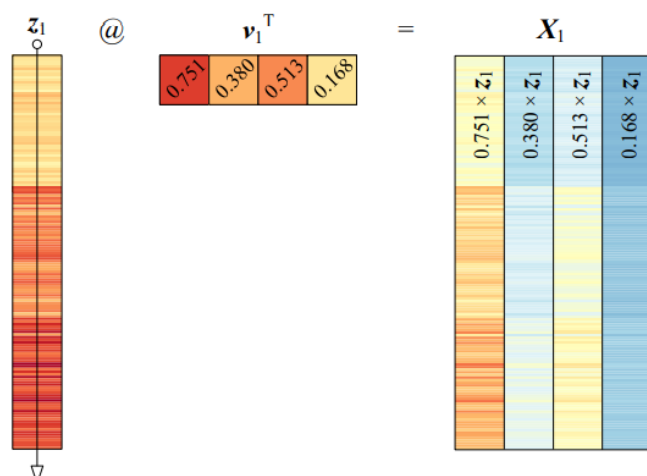
$$V = [v_1 \quad v_2 \quad v_3 \quad v_4] = \begin{bmatrix} 0.751 & 0.284 & 0.502 & 0.321 \\ 0.380 & 0.547 & -0.675 & -0.317 \\ 0.513 & -0.709 & -0.059 & -0.481 \\ 0.168 & -0.344 & -0.537 & 0.752 \end{bmatrix}$$

那么，我们先使用第一个基看一下：

- 首次投影：各个列以正交基的分量为系数，进行线性组合

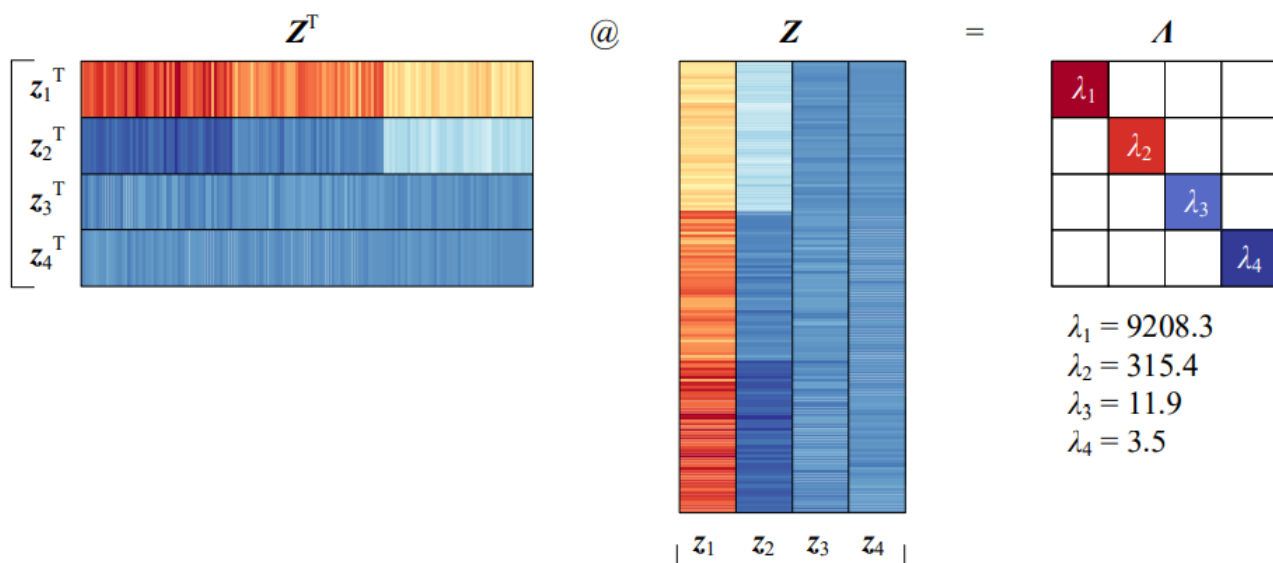


- 二次投影：生成向量投影——将这个向量乘以不同的系数变成各列



数据的正交化；主成分分析

我们考虑计算上面给出的 Z 的 Gram 矩阵，容易发现，这个矩阵是对角阵。这说明了我们选择的 V 矩阵的特殊性



注：这里有个小结论： X 的 Gram 矩阵的 trace 和 Z 的一样。
 由于

$$Z^T Z = V^T X^T X V = V^T G_X V = \Lambda$$

因此

$$G_X = V \Lambda V^T$$

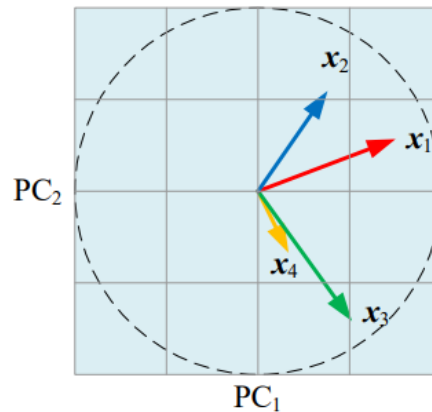
也就是说， X 的 Gram 矩阵实际上可以分解成三个矩阵的乘积。我们只要计算 X 的 Gram 矩阵如何对角化，就可以计算出 V^T 。

显然，我们每次得到的 Z 中的每一列元素都是由 X 中的列向量做线性组合得到的，而计算每一列的向量投影，再统统叠加，就又回到了 X 。我们可以写下 Z 中的每一列是如何组合的：

$$\begin{aligned} z_1 &= Xv_1 = 0.751x_1 + 0.380x_2 + 0.513x_3 + 0.168x_4 \\ z_2 &= Xv_2 = 0.284x_1 + 0.547x_2 - 0.709x_3 - 0.344x_4 \\ z_3 &= Xv_3 = 0.502x_1 - 0.675x_2 - 0.059x_3 - 0.537x_4 \\ z_4 &= Xv_4 = 0.321x_1 - 0.317x_2 - 0.481x_3 + 0.752x_4 \end{aligned}$$

$$V = \begin{bmatrix} 0.751 & 0.284 & 0.502 & 0.321 \\ 0.380 & 0.547 & -0.675 & -0.317 \\ 0.513 & -0.709 & -0.059 & -0.481 \\ 0.168 & -0.344 & -0.537 & 0.752 \end{bmatrix}$$

这里的 z_i ，就称为第 i 主成分。我们可以使用双标图 (biplot) 表示 X 的每一列对每个主成分的贡献，例如：



这几个主成分，实际上提取了数据中最主要的几个特征。例如，一首音乐可能有许多信息，但最主要的几个主成分可能是歌曲的速度、歌词的情感；一个高中生可能学习 9 门（甚至更多）课程，但最主要的两个主成分一定会反映该学生在文科和理科上的水平。

总结一下，主成分分析是一种特殊的投影—— $Z = XV$ ，其中， V 是一个特殊的正交矩阵。这样的投影后， Z 的各列是由 X 的各列线性组合而来，且互相正交。主成分分析提取了数据的主要变化方向。至于如何计算 V ，且听下回分解。