在讨论pca之前，我想先要阐述一下我对线性变换的理解。对于WX这种式子，我们通常会说它对X实施了线性变换，其中W为变换矩阵，X为向量。例如：

在这里，我想从几何角度上来解释线性变换。对于一个d维空间，任何一个向量都可以由d个线性无关向量所表示。所以当向量的位置改变了，我们可以理解为其基向量的位置改变了。因此，如果我们能够找到变换后空间的基向量，那么我们就能够计算原空间中的向量其所对应的目标向量。由于这是一个线性变换的过程，所以变换后的向量它仍保留着原来的线性关系。具体地，我们可以将其理解为：

原空间中的线性关系：

变换后的线性关系：

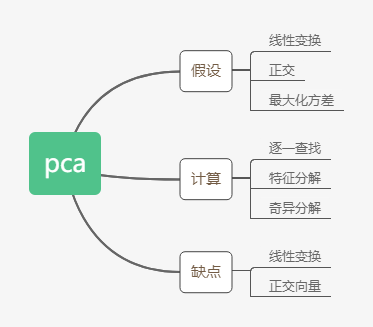
其中A为线性变换。

所以，我们可以将变换矩阵W中的每一列看作为新空间中的基向量。所以WX就是对新空间中的新基量进行线性变换。这种解释是我从bilibili网站上看来的。接下来，我想说一说我的理解。我个人认为这种理解可以更好地帮助我们理解非方阵的线性变换。在这里，我们所讨论的向量都是以原点为起始点的向量。所以，一般地我们只要确认了向量终点的坐标就能够确定向量的位置。在此，之前我们先看一下向量的内积：

该式子也会等于向量a在向量b上的投影乘以向量b的长度。在标准的坐标系中，

等式左边中向量的每一个坐标是向量与基向量的内积，由于标准基向量的长度为1。所以，向量与标准基向量的内积为向量在基向量方向上的投影长度。所以，对于线性变换，我们可以将行向量作为新空间中的基向量，而矩阵与向量相乘就是在计算向量与每一个基向量的内积。这样，我们就能够得到向量在新的空间中的坐标。所以对于降维变换，我们可以将其理解为其在原来的空间中找到某一个低维子空间，降维的过程则是计算原空间向量与低维子空间基向量的内积，这样我们就能够得到向量在低维子空间中的坐标。

对于给定的问题，我们会通过特征工程来选择一定的feature对其进行研究。通过特征工程，我们便创造以feature为为基向量的特征空间。但是，由于我们并不深知问题与feature或者是feature与feature之间的关系，通过特征工程一开始选择的feature并不一定是很好的feature。即feature之间可能存在者相关性，冗余性。所以，我们应该进行进一步的操作，在特征空间中找到一组更好地基向量来表示data。首先，data在该feature上的方差要最大；其次为了简化查找过程，基向量应该相互垂直。为什么一个好的feature是data在该feature上的方差最大呢？我们可以这样理解，假如data在feature上的方差很小，那么数据就缩成了一小团或者是一点。那么data在该feature上的value很有可能为常数，所以其失去了研究的意义。此外，data在feature上的方差较大，说明data在该feature上具有可分性。Pca正是一种帮助我们找到特征空间中方差最大的feature，以此来发现数据的潜在结构等。



在寻找主成分中，特征值或奇异值为data在新feature的方差。而且一般方差会集中于某些新feature之中。所以，我们可以挑选方差较大的feature来形成一个新的特征空间，以此达到对数据的降维。以前d个特征向量作为低维子空间的基向量，并将其标定为“新”的特征。“新”的特征可能是由已有的特征线性变换得到也有可能是已有特征中的某一个。所以，

其中W为投影矩阵，Z为X在投影空间中的坐标。

Note:一说到pca，我们的脑子里可能第一反应会想到降维。其实在我看来，pca还有另外一种好处，它可以将线性相关特征转换为线性无关特征。因为，对于对称矩阵，其必然存在着特征矩阵和特征向量。而且特征向量是线性无关的。所以我认为其可以将线性相关特征转换为线性无关特征。

Note:对于矩阵，我们可以从两个角度去理解它：几何和代数（线性方程组）