协方差矩阵的几何解释

内容[隐藏] [隐藏]

介绍

在本文中，我们通过探索线性变换与结果数据协方差之间的关系，提供协方差矩阵的直观，几何解释。大多数教科书基于协方差矩阵的概念来解释数据的形状。相反，我们采用向后的方法，并基于数据的形状解释协方差矩阵的概念。

在之前的一篇文章中，我们讨论了方差的概念，并提供了众所周知的公式来推导和证明，以估计样本方差。本文中使用图1来表明标准偏差（作为方差的平方根）可以衡量数据在特征空间中的分布程度。

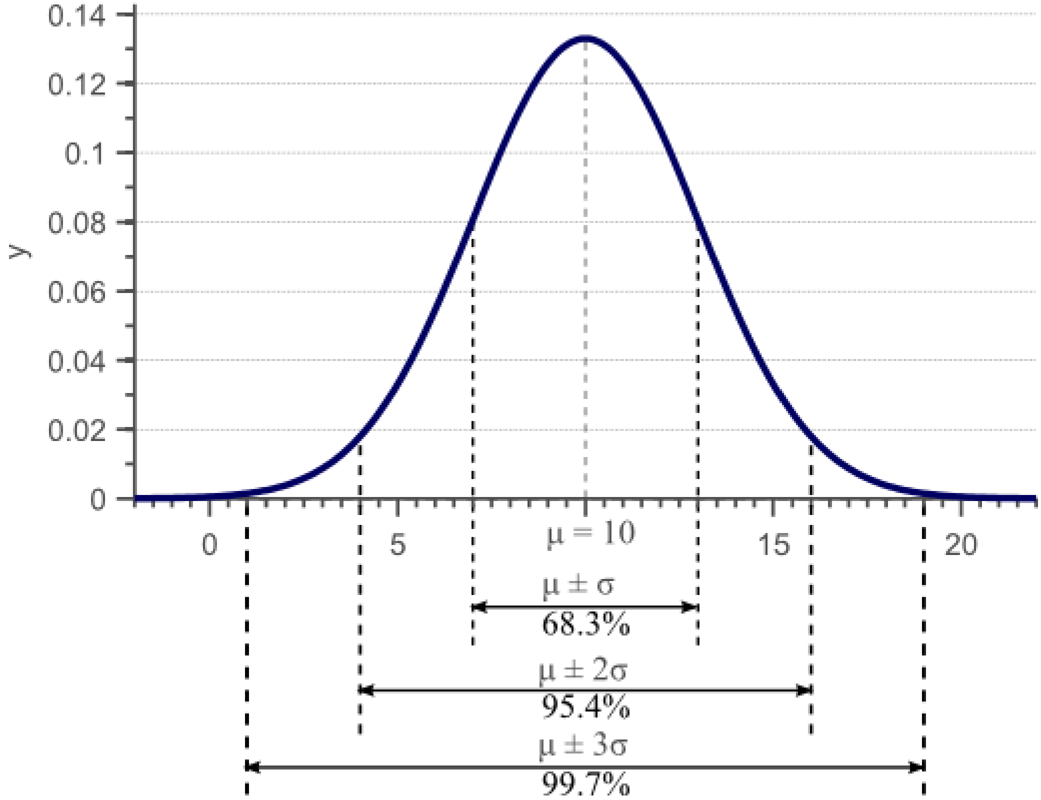


图1.高斯密度函数。对于正态分布数据，68%的样本落在由平均值加上和减去标准差定义的区间内。

我们发现样本方差的无偏估计可以通过以下方式获得：

$$\begin{aligned}\sigma_x^2 &= \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu)^2 \\ &= \mathbb{E}[(x-\mathbb{E}(x))(x-\mathbb{E}(x))] \\ &= \sigma(x,x)\end{aligned} \tag{1}$$

但是，方差只能用于解释数据在平行于特征空间轴的方向上的扩展。考虑图2所示的2D特征空间：

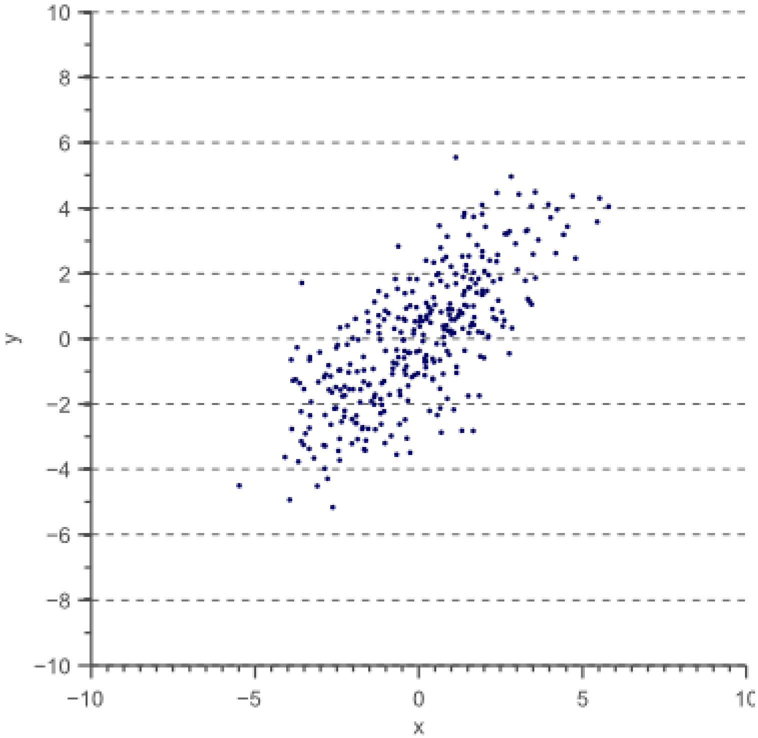


图2.协方差捕获数据的诊断传播。

对于这些数据，我们可以计算$\sigma(x, x)$x方向的方差$\sigma(y, y)$和y方向的方差。然而，数据的水平扩展和垂直扩展并不能解释清晰的对角线相关性。图2清楚地显示，平均而言，如果数据点的x值增加，则y值也增加，导致正相关。通过将方差概念扩展到所谓的数据的“协方差”，可以捕获这种相关性：

$$\sigma(x, y) = \mathbb{E}[(x - \mathbb{E}(x))(y - \mathbb{E}(y))] \tag{2}$$

对于二维数据，我们因此得到$\sigma(x, x)$，$\sigma(y, y)$，$\sigma(x, y)$和$\sigma(y, x)$。这四个值可以用矩阵表示，称为协方差矩阵：

$$\Sigma = \begin{bmatrix} \sigma(x, x) & \sigma(x, y) \\ \sigma(y, x) & \sigma(y, y) \end{bmatrix} \tag{3}$$

如果x与y正相关，则y也与x正相关。换句话说，我们可以说明这一点$\sigma(x, y) = \sigma(y, x)$。因此，协方差矩阵始终是对称矩阵，其对角线的方差和非对角线的协方差。二维正态分布数据完全由其均值和2×2协方差矩阵解释。类似地，3×3协方差矩阵用于捕获三维数据$N \times N$的扩展，协方差矩阵捕获N维数据的扩展。

图3说明了数据的整体形状如何定义协方差矩阵：

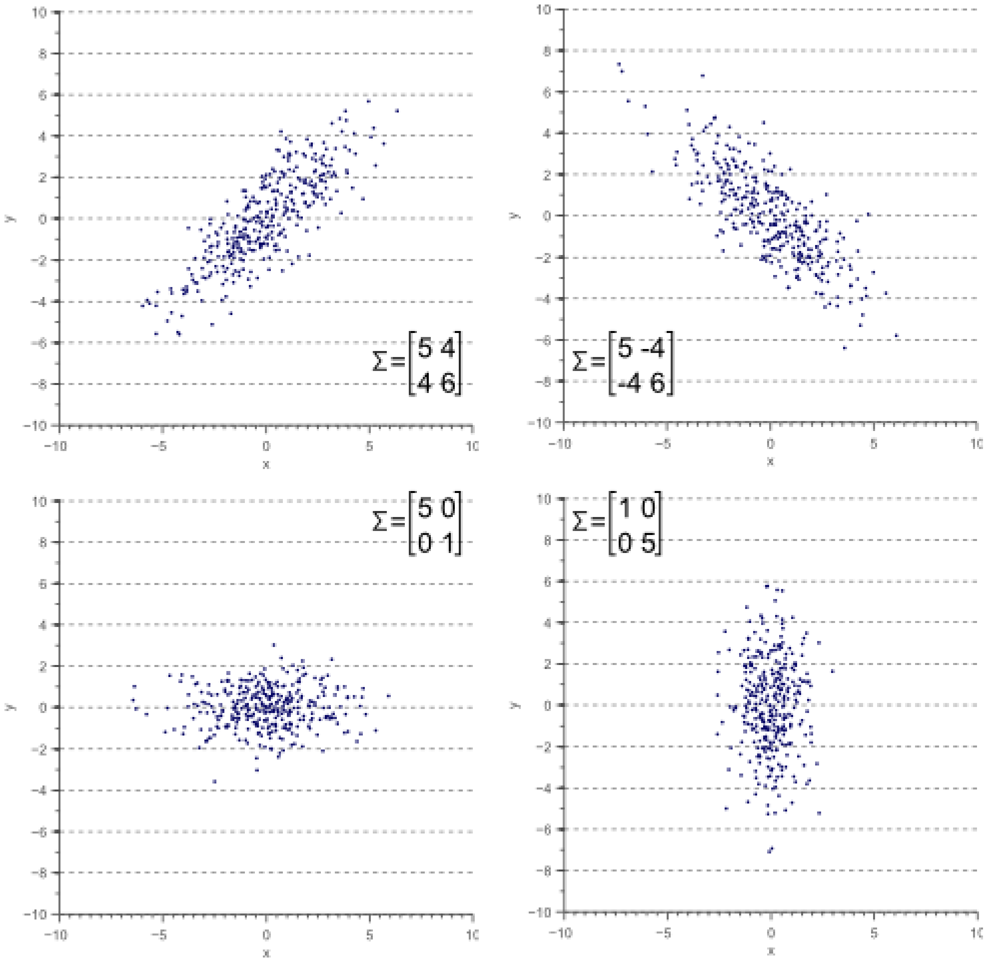


图3.协方差矩阵定义数据的形状。通过协方差捕获对角线扩展，而通过方差捕获轴对齐的扩展。

协方差矩阵的特征分解

在下一节中，我们将讨论如何将协方差矩阵解释为将白数据转换为我们观察到的数据的线性算子。但是，在深入研究技术细节之前，重要的是要直观地了解特征向量和特征值如何唯一地定义协方差矩阵，从而确定数据的形状。

正如我们在图3中看到的，协方差矩阵定义了数据的扩散（方差）和方向（协方差）。因此，如果我们想用一个向量及其大小来表示协方差矩阵，我们应该简单地尝试找到指向数据最大扩散方向的向量，其大小等于此中的扩散（方差）。方向。

如果我们将此向量定义为$\vec{v}$，则将我们的数据D投影到此向量上$\vec{v}^\intercal D$，并获得投影数据的方差$\vec{v}^\intercal \Sigma \vec{v}$。由于我们正在寻找$\vec{v}$指向最大方差方向的向量，我们应该选择其组件，使得$\vec{v}^\intercal \Sigma \vec{v}$投影数据的协方差矩阵尽可能大。最大化$\vec{v}^\intercal \Sigma \vec{v}$关于形式的任何函数$\vec{v}$，其中$\vec{v}$标准化单位向量可以被表达为所谓的瑞利商（Rayleigh Quotient）。通过设置$\vec{v}$等于矩阵的最大特征向量来获得这种瑞利商的最大值Σ。

换句话说，协方差矩阵的最大特征向量总是指向数据的最大方差的方向，并且该向量的大小等于对应的特征值。第二大特征向量总是与最大特征向量正交，并指向数据的第二大扩展方向。

现在让我们来看看一些例子。在前面的文章中，我们看到线性变换矩阵T完全由其特征向量和特征值定义。应用于协方差矩阵，这意味着：

$$\Sigma \vec{v} = \lambda \vec{v} \tag{4}$$

其中$\vec{v}$是特征向量Σ，并且λ是相应的特征值。

如果我们的数据的协方差矩阵是对角矩阵，使得协方差为零，那么这意味着方差必须等于特征值λ。这由图4示出，其中特征向量以绿色和品红色示出，并且其中特征值明显等于协方差矩阵的方差分量。

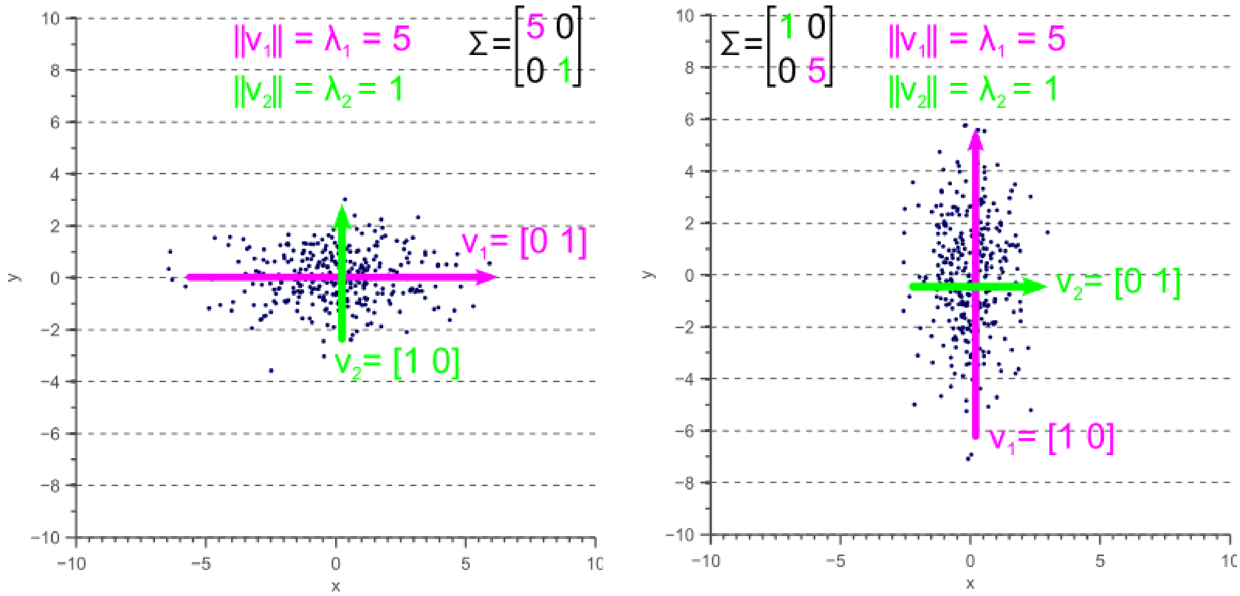


图4.协方差矩阵的特征向量

然而，如果协方差矩阵不是对角线的，使得协方差不为零，则情况稍微复杂一些。特征值仍然表示数据的最大扩展方向上的方差幅度，并且协方差矩阵的方差分量仍然表示x轴和y轴方向上的方差幅度。但由于数据不是轴对齐的，因此这些值不再相同，如图5所示。

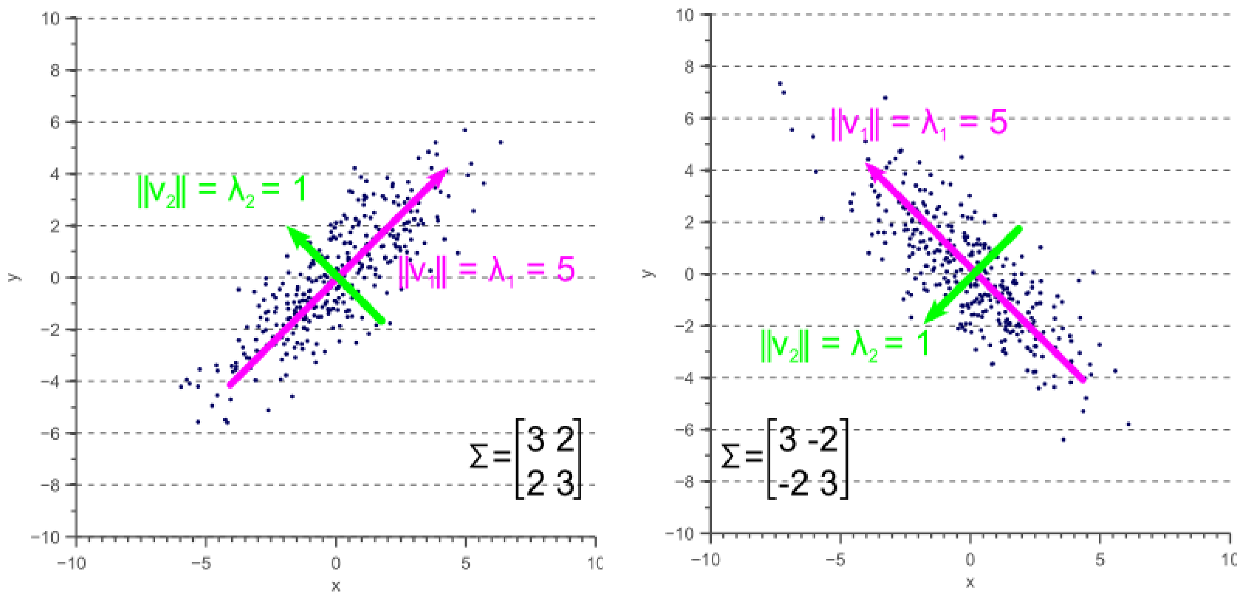


图5.特征值与方差的关系

通过将图5与图4进行比较，可以清楚地看出特征值表示沿着特征向量方向的数据的方差，而协方差矩阵的方差分量表示沿着轴的扩展。如果没有协方差，则两个值都相等。

协方差矩阵作为线性变换

现在让我们暂时忘掉协方差矩阵。图3中的每个示例都可以简单地视为图6的线性变换实例：

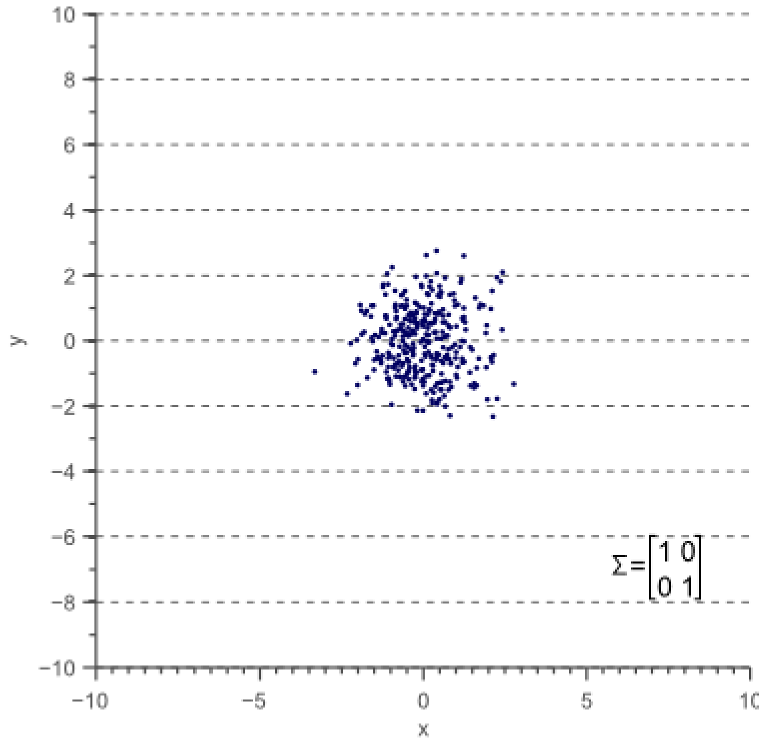


图6.具有单位协方差矩阵的数据称为白色数据。

假设图6所示的数据D，则可以通过线性变换获得图3所示的每个示例D：

$$D' = T\,D \tag{5}$$

其中T是由旋转矩阵R和缩放矩阵组成的变换矩阵S：

$$T = R\,S. \tag{6}$$

这些矩阵定义为：

$$R = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \tag{7}$$

θ旋转角度在哪里，并且：

$$S = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \tag{8}$$

其中s_x和s_y是分别在x方向和y方向上的缩放因子。

在下面的段落中，我们将讨论协方差矩阵Σ和线性变换矩阵之间的关系$T = R\,S$。

让我们从未缩放（缩放等于1）和未旋转数据开始。在统计中，这通常被称为“白色数据”，因为它的样本是从标准正态分布中提取的，因此对应于白色（不相关）噪声：

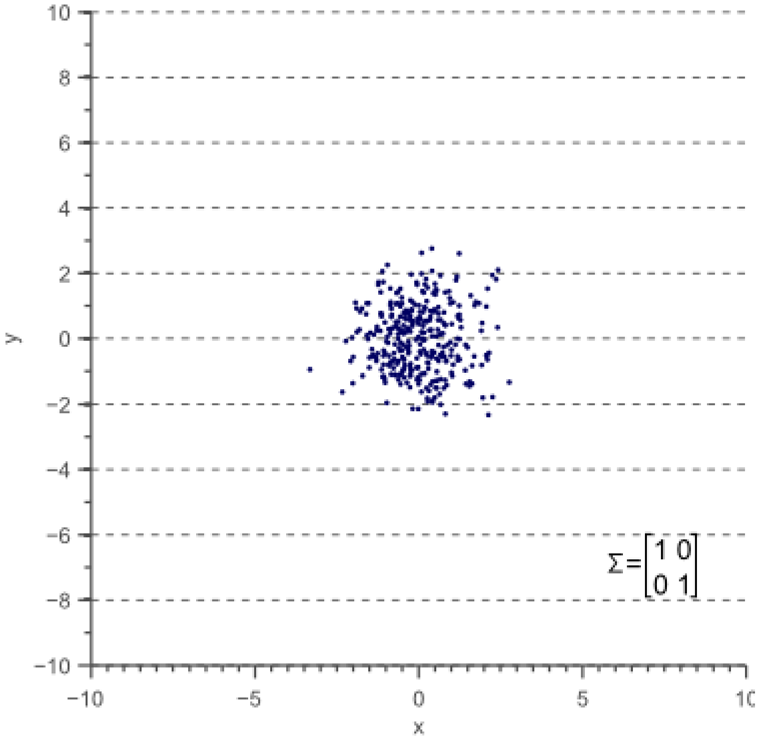


图7.白色数据是具有单位协方差矩阵的数据。

该“白色”数据的协方差矩阵等于单位矩阵，使得方差和标准差等于1且协方差等于零：

$$\Sigma = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{9}$$

现在让我们使用因子4在x方向上缩放数据：

$$D' = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} D \tag{10}$$

D'现在的数据如下：

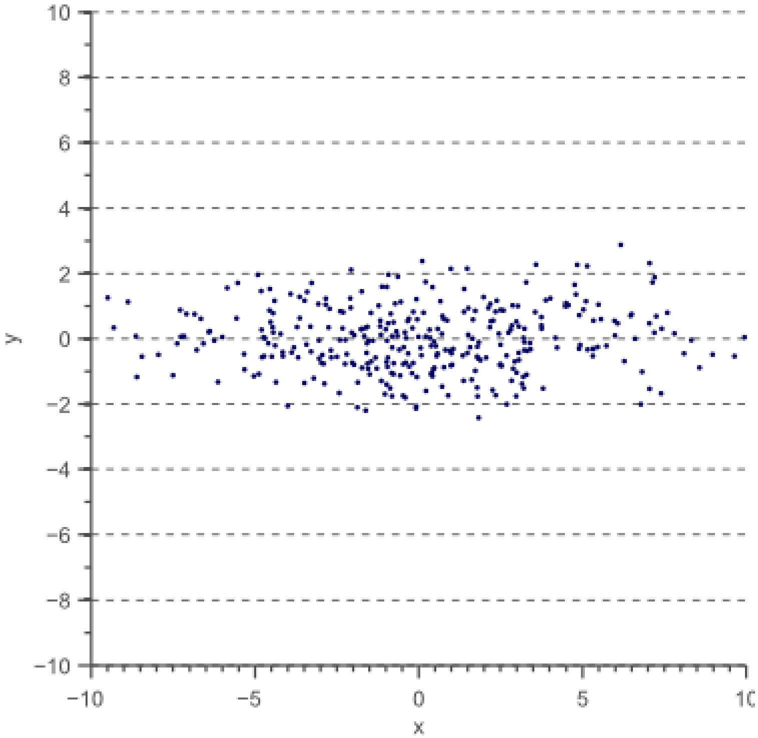


图8. x方向的方差导致水平缩放。

协方差矩阵Σ'的D'现在是：

$$\Sigma' = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix} = \begin{bmatrix} 16 & 0 \\ 0 & 1 \end{bmatrix} \tag{11}$$

因此，协方差矩阵Σ'所得到的数据的D'是相关的线性变换T被施加到如下的原始数据：$D' = T\,D$，其中

$$T = \sqrt{\Sigma'} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}. \tag{12}$$

然而，尽管当数据在x和y方向上缩放时等式（12）成立，但是如果在应用旋转时它也成立则问题上升。为了研究一般情况T下线性变换矩阵与协方差矩阵之间的关系Σ'，我们将尝试将协方差矩阵分解为旋转矩阵和缩放矩阵的乘积。

如前所述，我们可以通过其特征向量和特征值来表示协方差矩阵：

$$\Sigma\vec{v} = \lambda\vec{v} \tag{13}$$

其中$\vec{v}$是特征向量Σ，并且λ是相应的特征值。

等式（13）适用于每个特征向量 - 特征值矩阵对Σ。在2D情况下，我们获得两个特征向量和两个特征值。由等式（13）定义的两个方程的系统可以使用矩阵表示法有效地表示：

$$\Sigma\,V = V\,L \tag{14}$$

其中V列是特征向量的矩阵在哪里，Σ并且L是对角矩阵，其非零元素是相应的特征值。

这意味着我们可以将协方差矩阵表示为其特征向量和特征值的函数：

$$\Sigma = V\, L\, V^{-1} \tag{15}$$

等式（15）被称为协方差矩阵的特征分解，并且可以使用奇异值分解算法来获得。尽管特征向量表示数据的最大方差的方向，但是特征值表示这些方向上的该方差的大小。换句话说，V表示旋转矩阵，而$\sqrt{L}$表示缩放矩阵。因此，协方差矩阵可以进一步分解为：

$$\Sigma = R\, S\, S\, R^{-1} \tag{16}$$

其中$R = V$是旋转矩阵，$S = \sqrt{L}$是一个缩放矩阵。

在等式（6）中，我们定义了线性变换$T = R\, S$。由于S是对角缩放矩阵，$S = S^{\intercal}$。此外，由于R是正交矩阵，$R^{-1} = R^{\intercal}$。因此，$T^{\intercal} = (R\, S)^{\intercal} = S^{\intercal}\, R^{\intercal} = S\, R^{-1}$。协方差矩阵因此可以写成：

$$\Sigma = R\, S\, S\, R^{-1} = T\, T^{\intercal}, \tag{17}$$

换句话说，如果我们将由图7所示$T = R\, S$的原始白色数据定义的线性变换应用D，我们获得D'具有协方差矩阵的旋转和缩放数据$T\, T^{\intercal} = \Sigma' = R\, S\, S\, R^{-1}$。如图10所示：

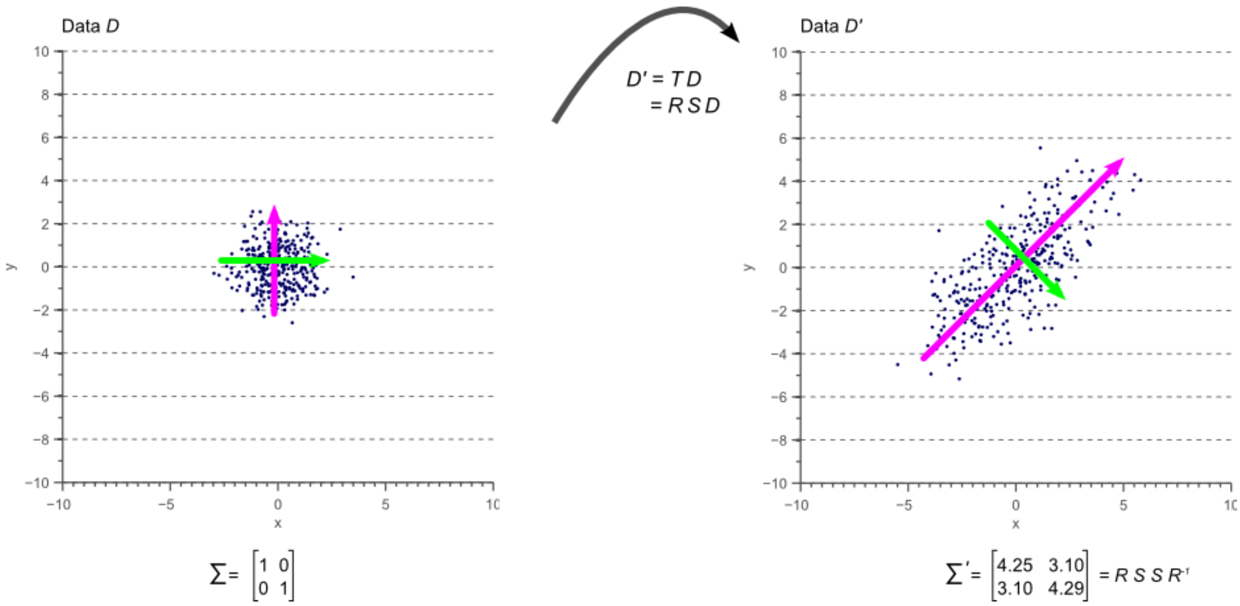


图10.协方差矩阵表示原始数据的线性变换。

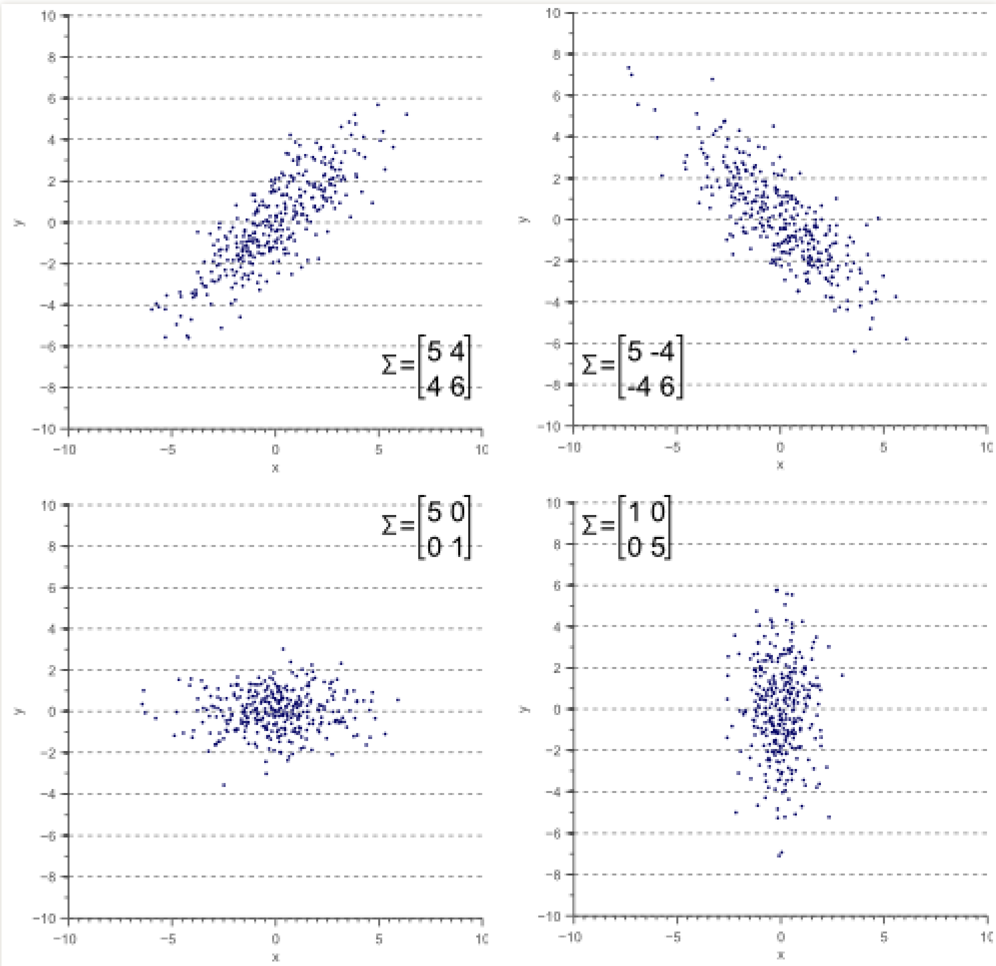
图10中的彩色箭头表示特征向量。最大的特征向量，即具有最大对应特征值的特征向量，总是指向数据的最大方差的方向，从而定义其方向。由于旋转矩阵的正交性，后续的特征向量总是与最大的特征向量正交。

结论

在本文中，我们展示了观测数据的协方差矩阵与白色不相关数据的线性变换直接相关。该线性变换完全由数据的特征向量和特征值定义。虽然特征向量表示旋转矩阵，但特征值对应于每个维度中的缩放因子的平方。

如果您是这个博客的新手，请不要忘记订阅，或者[在Twitter上关注我](#)！

摘要



文章名	协方差矩阵的几何解释
作者	文森特斯普鲁特

描述	在本文中，我们通过探索线性变换与结果数据协方差之间的关系，提供协方差矩阵的直观，几何解释。

与您的社交网络分享这篇文章：

分享

2014年4月24日 文森特Spruyt 线性代数 47条评论 协方差矩阵，特征值分解，特征向量，线性变换，PCA

« 分类中的维度诅咒 使用PCA进行特征提取 »

评论

克里斯 说：

2014年5月14日下午3:42

好文章谢谢

答复

亚历克斯 说：

2014年5月14日晚上8:09

协方差矩阵是对称的。因此，我们可以找到正交特征向量的基础，然后是$ \ Sigma = VL V ^ T $。
从计算的角度来看，找到$ V ^ T $要比$ V ^ { - 1} $简单得多。

答复