Go to... ▼

A geometric interpretation of the covariance matrix

Contents [hide]

Introduction

In this article, we provide an intuitive, geometric interpretation of the covariance matrix, by exploring the relation between linear transformations and the resulting data covariance. Most textbooks explain the shape of data based on the concept of covariance matrices. Instead, we take a backwards approach and explain the concept of covariance matrices based on the shape of data.

In a previous article, we discussed the concept of variance, and provided a derivation and proof of the well known formula to estimate the sample variance. Figure 1 was used in this article to show that the standard deviation, as the square root of the variance, provides a measure of how much the data is spread across the feature space.
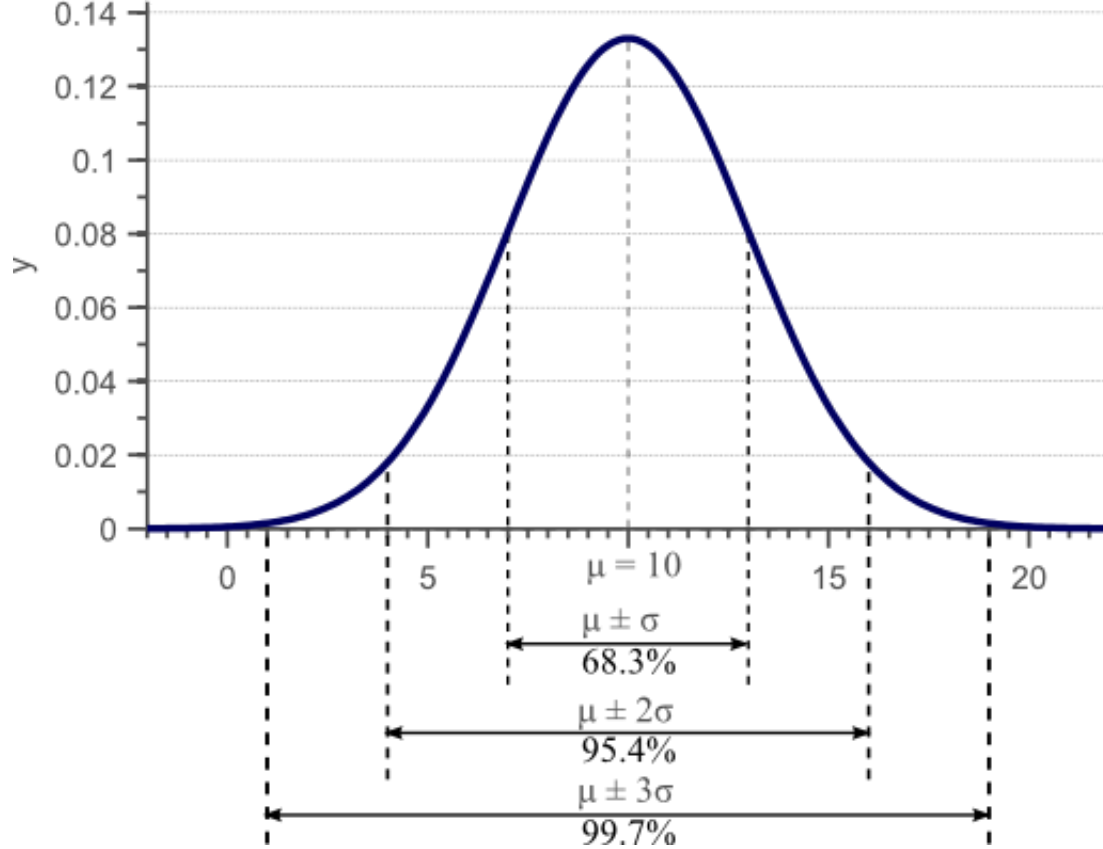
Figure 1. Gaussian density function. For normally distributed data, 68% of the samples fall within the interval defined by the mean plus and minus the standard deviation.

We showed that an unbiased estimator of the sample variance can be obtained by:

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

However, variance can only be used to explain the spread of the data in the directions parallel to the axes of the feature space. Consider the 2D feature space shown by figure 2:
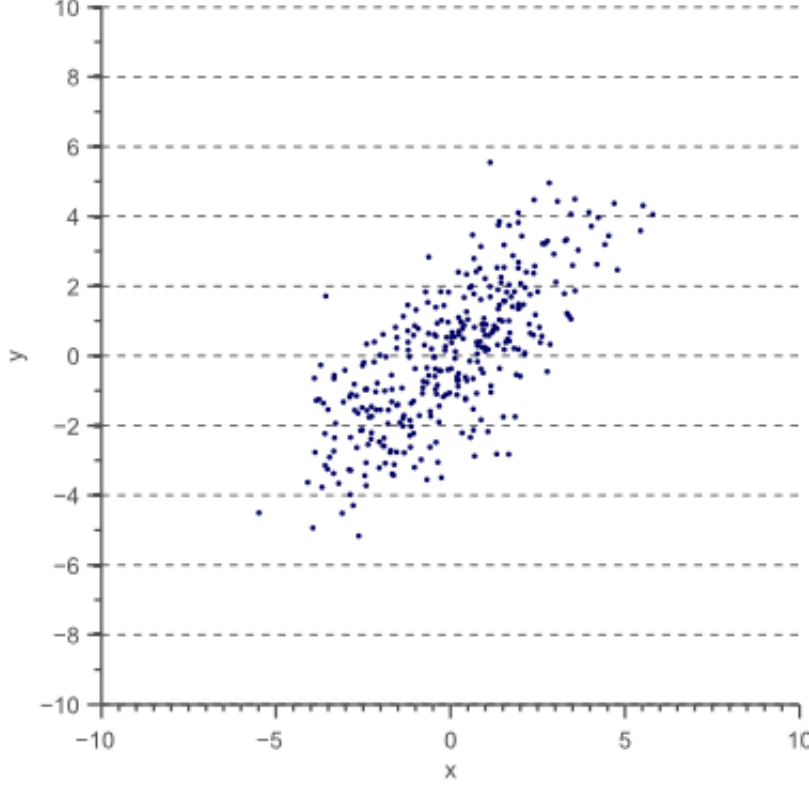
Figure 2. The diagnoal spread of the data is captured by the covariance.

For this data, we could calculate the variance $\sigma(x, x)$ in the x-direction and the variance $\sigma(y, y)$ in the y-direction. However, the horizontal spread and the vertical spread of the data does not explain the clear diagonal correlation. Figure 2 clearly shows that on average, if the x-value of a data point increases, then also the y-value increases, resulting in a positive correlation. This correlation can be captured by extending the notion of variance to what is called the 'covariance' of the data:
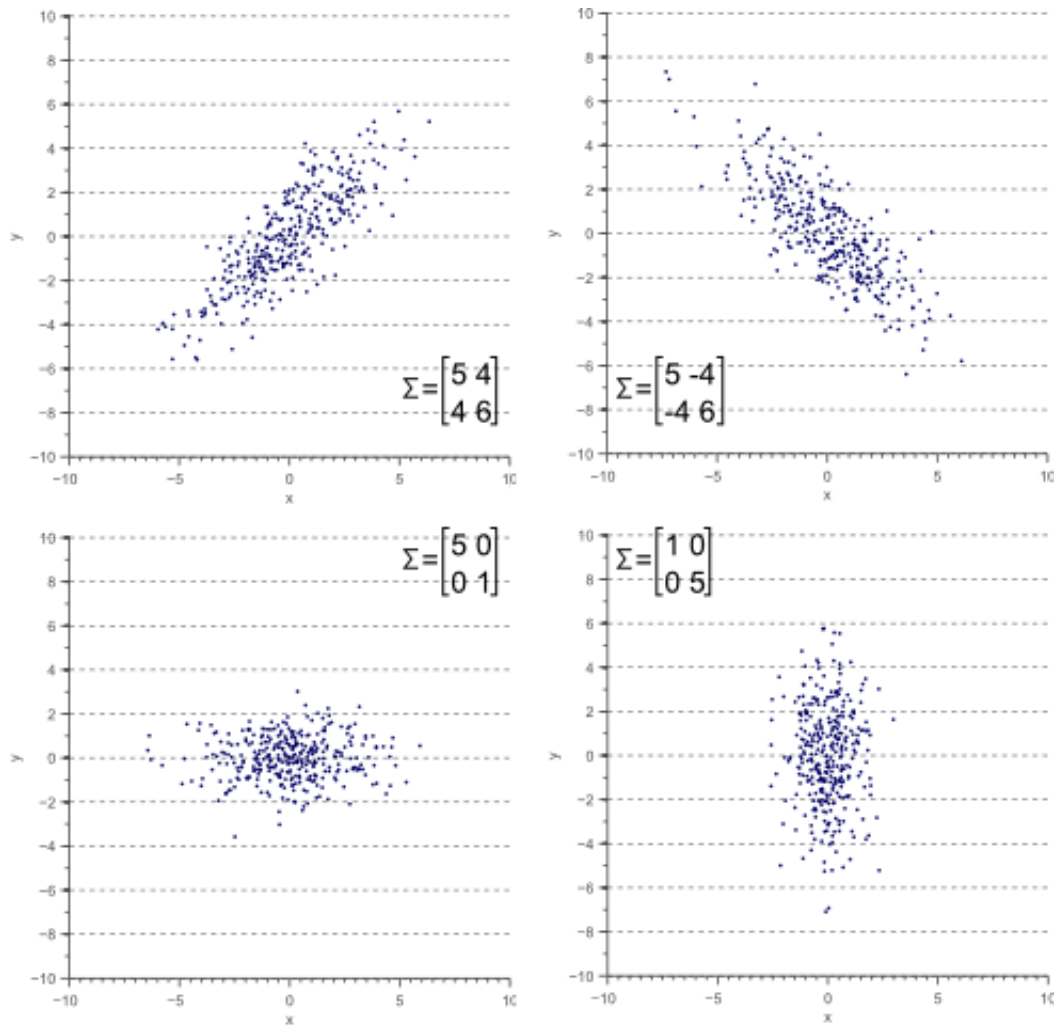
$$\sigma(x, y) = \mathbb{E}[(x - \mathbb{E}(x))(y - \mathbb{E}(y))] \tag{2}$$

For 2D data, we thus obtain $\sigma(x, x), \sigma(y, y), \sigma(x, y)$ and $\sigma(y, x)$. These four values can be summarized in a matrix, called the covariance matrix:

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

If x is positively correlated with y, y is also positively correlated with x. In other words, we can state that $\sigma(x, y) = \sigma(y, x)$. Therefore, the covariance matrix is always a symmetric matrix with the variances on its diagonal and the covariances off-diagonal. Two-dimensional normally distributed data is explained completely by its mean and its 2×2 covariance matrix. Similarly, a 3×3 covariance matrix is used to capture the spread of three-dimensional data, and a $N \times N$ covariance matrix captures the spread of N-dimensional data.

Figure 3 illustrates how the overall shape of the data defines the covariance matrix:

$$\Sigma = \begin{bmatrix} 5 & 4 \\ 4 & 6 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 5 & -4 \\ -4 & 6 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

Figure 3. The covariance matrix defines the shape of the data. Diagonal spread is captured by the covariance, while axis-aligned spread is captured by the variance.

Now let's forget about covariance matrices for a moment. Each of the examples in figure 3 can simply be considered to be a linearly transformed instance of figure 4:

$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
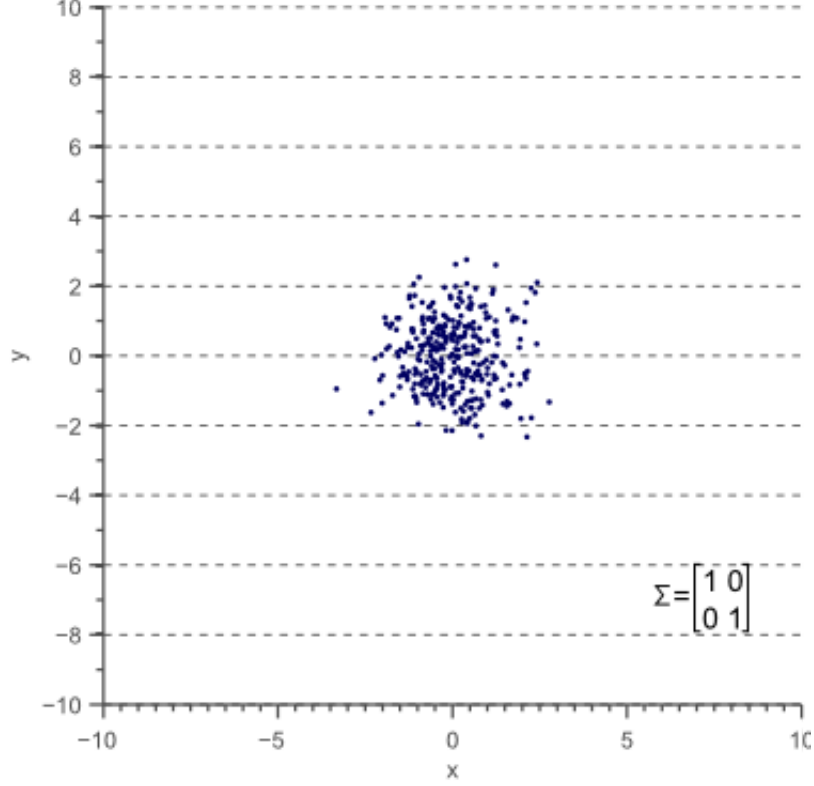
Figure 4. Data with unit covariance matrix is called white data.

Let the data shown by figure 4 be D, then each of the examples shown by figure 3 can be obtained by linearly transforming D:

$$D' = T D \tag{4}$$

where T is a transformation matrix consisting of a rotation matrix R and a scaling matrix S:

$$T = R S. \tag{5}$$

These matrices are defined as:

$$R = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \tag{6}$$

where θ is the rotation angle, and:

$$S = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \tag{7}$$

where S_x and S_y are the scaling factors in the x direction and the y direction respectively.

In the following section, we will discuss the relation between the covariance matrix Σ, and the linear transformation matrix $T = R\,S$.

Covariance matrix as a linear transformation

Let's start with unscaled (scale equals 1) and unrotated data. In statistics this is often refered to as 'white data' because its samples are drawn from a standard normal distribution and therefore correspond to white (uncorrelated) noise:
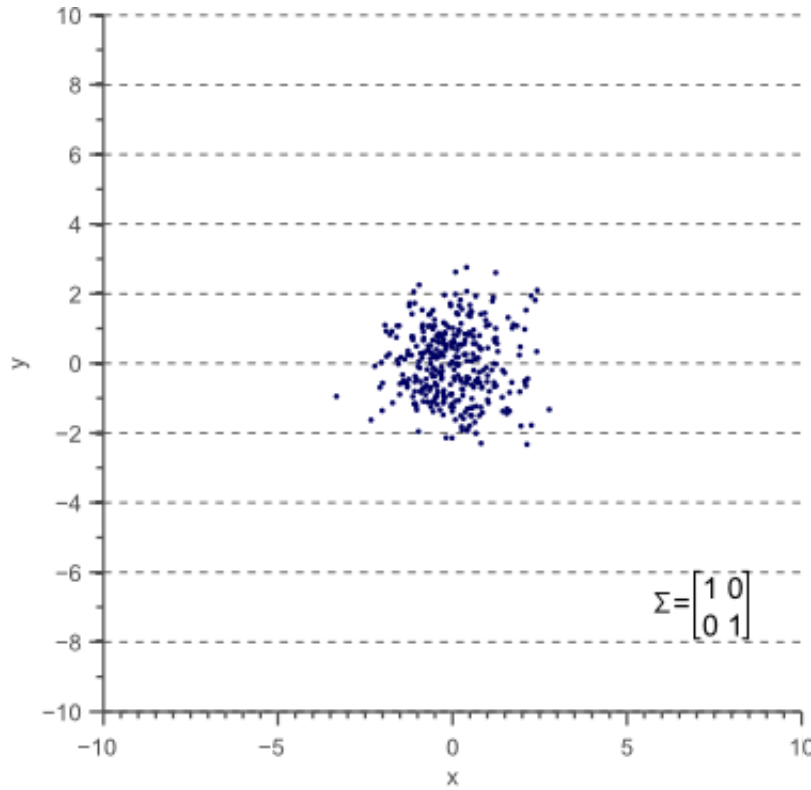
Figure 5. White data is data with a unit covariance matrix.
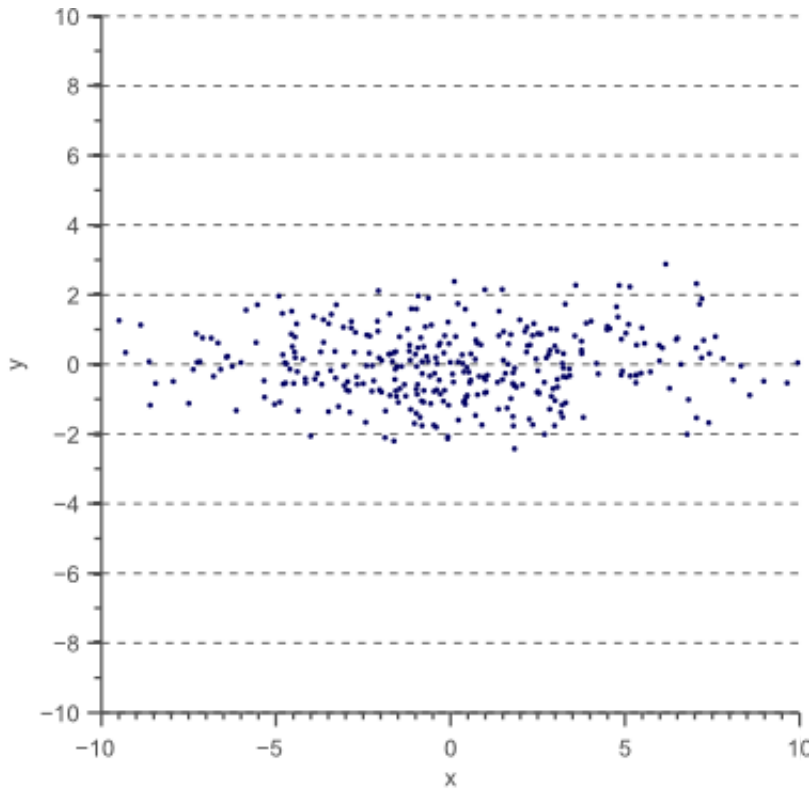
The covariance matrix of this 'white' data equals the identity matrix, such that the variances and standard deviations equal 1 and the covariance equals zero:

$$\Sigma = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{8}$$

Now let's scale the data in the x-direction with a factor 4:

$$D' = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} D \tag{9}$$

The data D' now looks as follows:

Figure 6. Variance in the x-direction results in a horizontal scaling.

The covariance matrix Σ' of D' is now:

$$\Sigma' = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix} = \begin{bmatrix} 16 & 0 \\ 0 & 1 \end{bmatrix} \tag{10}$$

Thus, the covariance matrix Σ' of the resulting data D' is related to the linear transformation T that is applied to the original data as follows: $D' = T\,D$, where

$$T = \sqrt{\Sigma'} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}. \tag{11}$$

However, although equation (11) holds when the data is scaled in the x and y direction, the question rises if it also holds when a rotation is applied. To investigate the relation between the linear transformation matrix T and the covariance matrix Σ' in the general case, we will therefore try to decompose the covariance matrix into the product of rotation and scaling matrices.

In an earlier article we saw that a linear transformation matrix T is completely defined by its eigenvectors and

eigenvalues. Applied to the covariance matrix, this means that:

$$\Sigma \vec{v} = \lambda \vec{v} \tag{12}$$

where $\vec{v}$ is an eigenvector of Σ, and λ is the corresponding eigenvalue.

Equation ([12](#)) holds for each eigenvector-eigenvalue pair of matrix Σ. In the 2D case, we obtain two eigenvectors and two eigenvalues. The system of two equations defined by equation ([12](#)) can be represented efficiently using matrix notation:

$$\Sigma V = V L \tag{13}$$

where V is the matrix whose columns are the eigenvectors of Σ and L is the diagonal matrix whose non-zero elements are the corresponding eigenvalues.

This means that we can represent the covariance matrix as a function of its eigenvectors and eigenvalues:

$$\Sigma = V L V^{-1} \tag{14}$$

Equation ([14](#)) is called the eigendecomposition of the covariance matrix and can be obtained using a Singular Value Decomposition algorithm. Whereas the eigenvectors represent the directions of the largest variance of the data, the eigenvalues represent the magnitude of this variance in those directions. In other words, V represents a rotation matrix, while $\sqrt{L}$ represents a scaling matrix. The covariance matrix can thus be decomposed further as:

$$\Sigma = R S S R^{-1} \tag{15}$$

where $R = V$ is a rotation matrix and $S = \sqrt{L}$ is a scaling matrix.

In equation ([5](#)) we defined a linear transformation $T = R S$. Since S is a diagonal scaling matrix, $S = S^{\mathsf{T}}$. Furthermore, since R is an orthogonal matrix, $R^{-1} = R^{\mathsf{T}}$. Therefore, $T^{\mathsf{T}} = (R S)^{\mathsf{T}} = S^{\mathsf{T}} R^{\mathsf{T}} = S R^{-1}$. The covariance matrix can thus be written as:

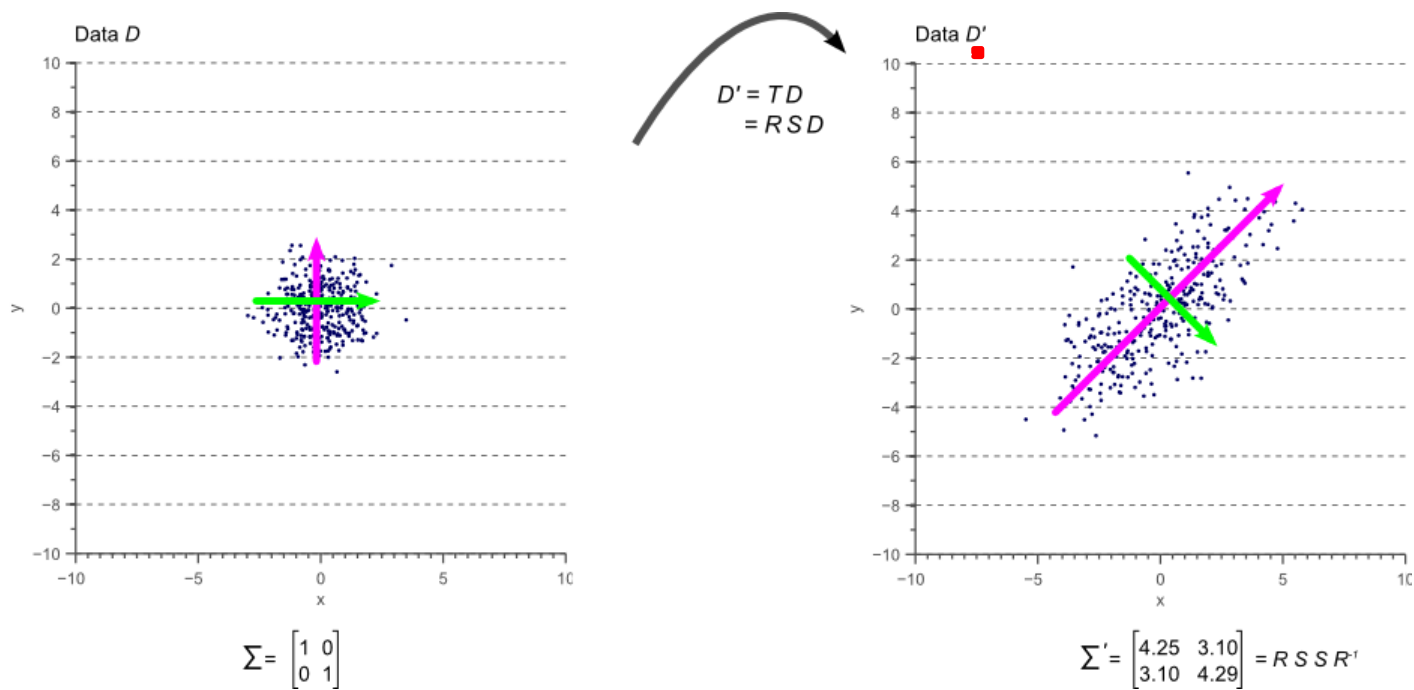$$\Sigma = R S S R^{-1} = T T^{\mathsf{T}} = T^2, \tag{16}$$

and thus

$$T = \sqrt{\Sigma}. \tag{17}$$

In other words, if we apply the linear transformation defined by $T = R S$ to the original white data D shown by figure 5, we obtain the rotated and scaled data D' with covariance matrix $\Sigma' = R S S R^{-1} = T^2$.

This is illustrated by figure 7:

Figure 7. The covariance matrix represents a linear transformation of the original data.

The colored arrows in figure 7 represent the eigenvectors. The largest eigenvector, i.e. the eigenvector with the largest corresponding eigenvalue, always points in the direction of the largest variance of the data and thereby defines its orientation. Subsequent eigenvectors are always orthogonal to the largest eigenvector due to the orthogonality of rotation matrices.

Conclusion

In this article we showed that the covariance matrix of observed data is directly related to a linear transformation of white, uncorrelated data. This linear transformation is completely defined by the eigenvectors and eigenvalues of the data. While the eigenvectors represent the rotation matrix, the eigenvalues correspond to the square of the scaling factor in each dimension.

Summary
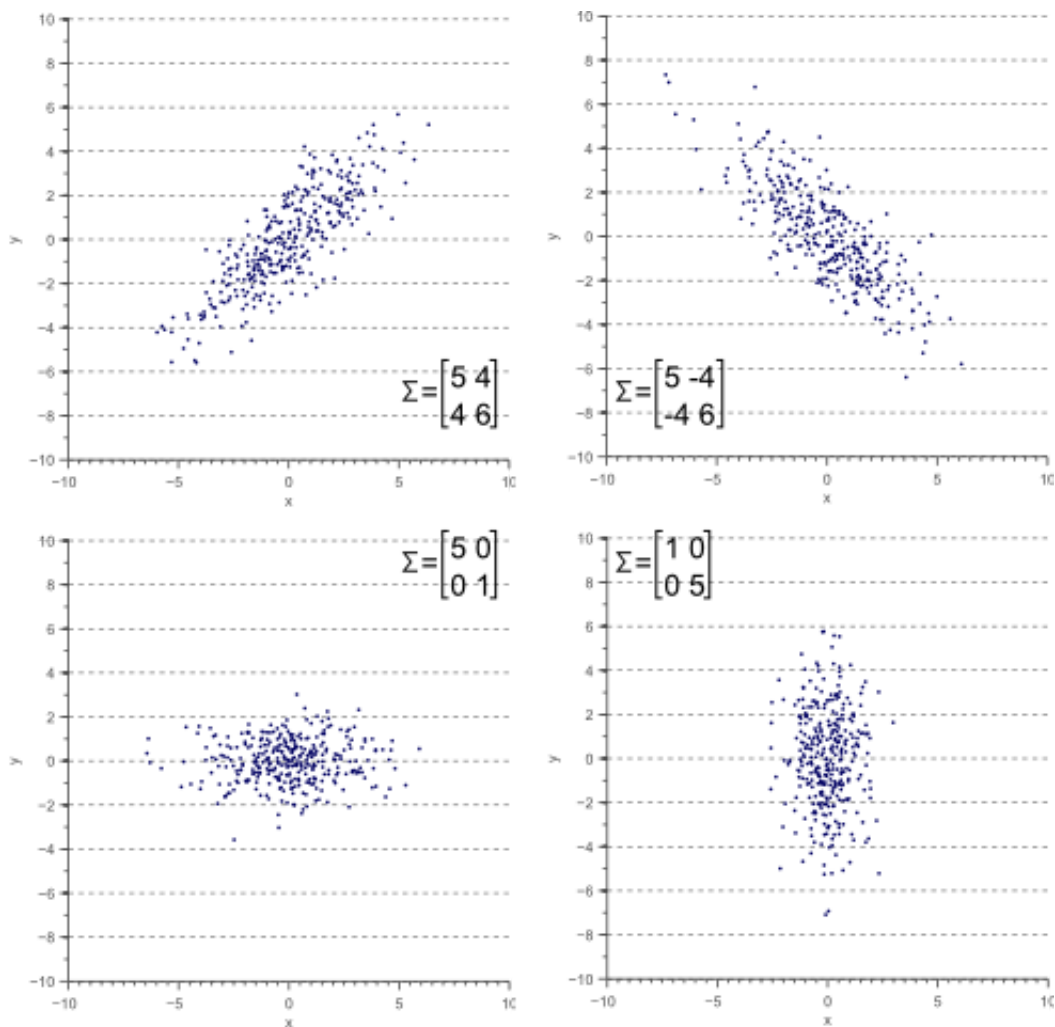
Article Name	A geometric interpretation of the covariance matrix
Author	Vincent Spruyt
Description	In this article, we provide an intuitive, geometric interpretation of the covariance matrix, by exploring the relation between linear transformations and the resulting data covariance.

Share this post with your social networks:

Share

April 24, 2014 Vincent Spruyt Linear algebra 8 Comments covariance matrix,

eigendecomposition, Eigenvectors, linear transformation, PCA

«The Curse of Dimensionality in classification

Feature extraction using PCA»

Comments

Chris says:

May 14, 2014 at 3:42 pm

Great article thank you

Reply

Alex says:

May 14, 2014 at 8:09 pm

The covariance matrix is symmetric. Hence we can find a basis of orthonormal eigenvectors and then $\Sigma=VL V^T$.
From computational point of view it is much simpler to find V^T than V^{-1}.

Reply

Vincent Spruyt says:

May 15, 2014 at 7:59 am

Very true, Alex, and thanks for your comment! This is also written in the article: "Furthermore, since R is an orthogonal matrix, R^{-1} = R^T". But you are right that I only mention this near the end of the

article, mostly because it is easier to develop an intuitive understanding of the first part of the article by considering R^{-1} instead of R^T.

Reply

Brian says:

May 24, 2014 at 1:03 am

Great post! I had a couple questions:

1) The data D doesn't need to be Gaussian does it?

2) Is [9] reversed (should D be on the left)?

Reply

Vincent Spruyt says:

May 24, 2014 at 8:27 am

Hi Brian:

1) Indeed the data D does not need to be Gaussian for the theory to hold, I should probably have made that more clear in the article. However, talking about covariance matrices often does not have much meaning in highly non-Gaussian data.

2) That depends on whether D is a row vector or a

column vector I suppose. In this case, if each
column of D is a data entry, then R*D = (D^t*R)^t

Gino says:

August 14, 2014 at 7:26 pm

Why is TT' = T^2 ?

Konstantin says:

August 18, 2014 at 1:47 am

Thank you for this great post! But let me please correct one
fundamental mistake that you made. The square root of
covariance matrix M is not equal to R * S. The square root of
M equals R * S * R', where R' is transposed R. Proof: (R * S *
R') * (R * S * R') = R * S * R' * R * S * R' = R * S * S * R' = T *
T' = M. And, of course, T is not a symmetric matrix (in your
post T = T', which is wrong).

Vincent Spruyt says:

August 18, 2014 at 8:29 am

Thanks a lot for noticing! You are right indeed, I will get back about this soon (don't really have time right now).

Reply

Comments Are Very Welcome!

Enter your comment here...

Search   Go

Article Topics

- Dimensionality reduction (2)
  - Feature extraction (2)
- Math basics (4)
  - Linear algebra (2)
  - Statistics (2)

Recent Posts

- Feature extraction using PCA

- A geometric interpretation of the covariance matrix
- The Curse of Dimensionality in classification
- How to draw a covariance error ellipse?
- Why divide the sample variance by N-1?

Follow This Blog

g+ Follow   142

Follow

Subscribe Via Email

Enter your email address to subscribe to this blog and receive notifications of new posts by email.

Email Address

Subscribe

Rss Feed

RSS - Posts

RSS - Comments