Eigenvectors and Eigenvalues

Probability and Statistical Inference

Bojan Božić

TU Dublin

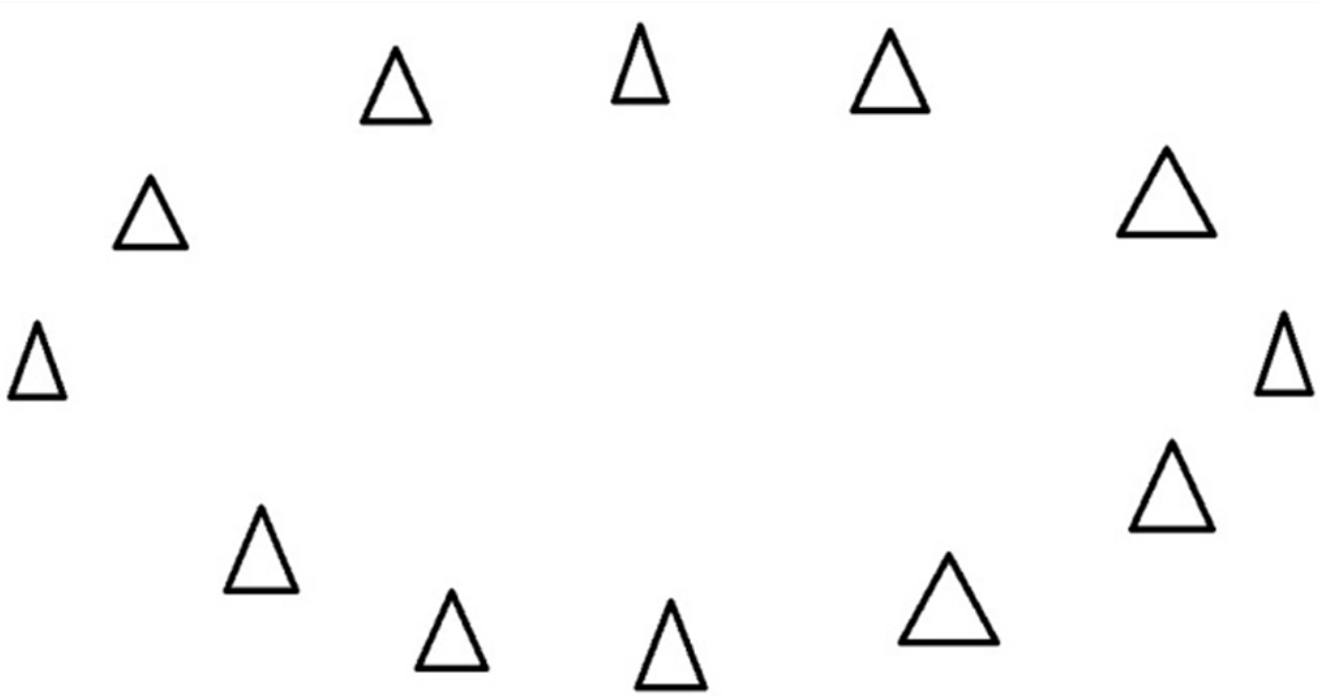
Winter 2020

Principal Components

- The underlying structure of the data.
- The directions where there is most variance.

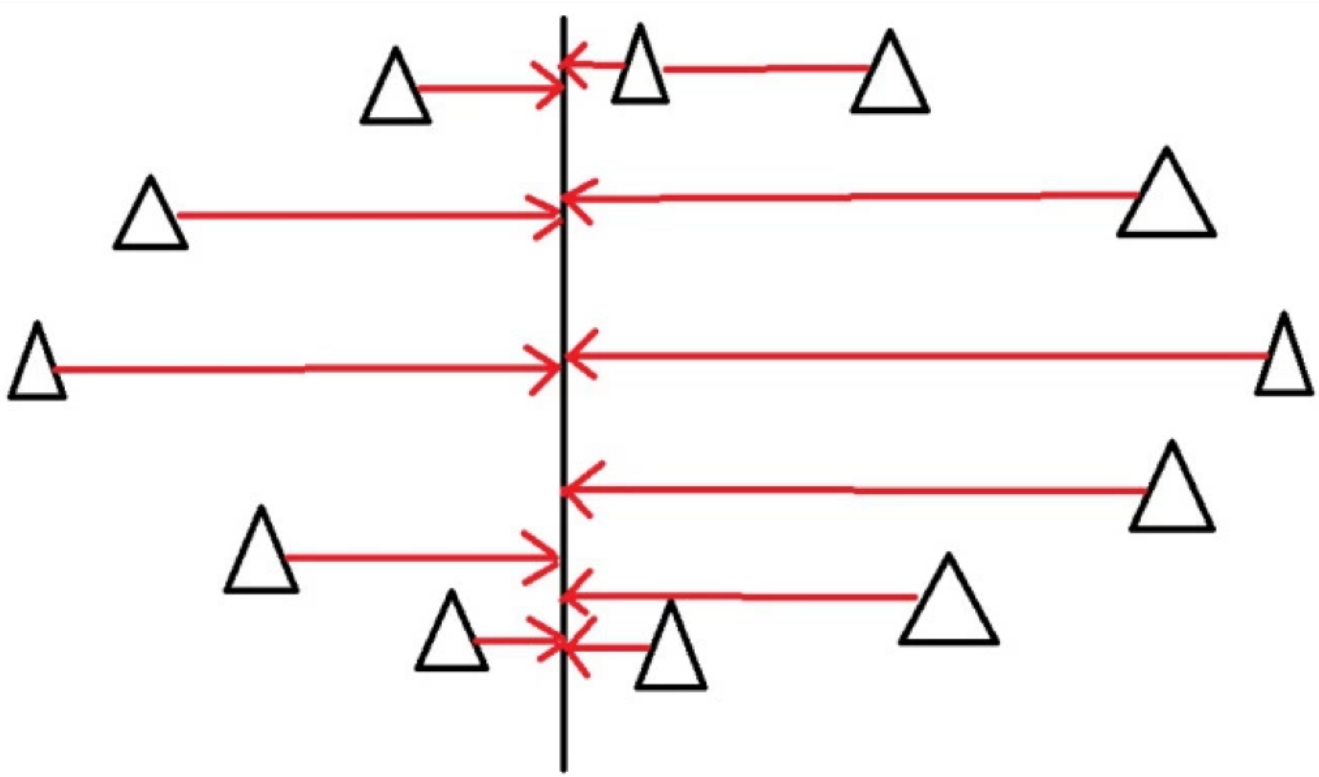
Example

The triangles are points of data. To find the direction find the straight line where the data is most spread out when it is projected onto it.



Example

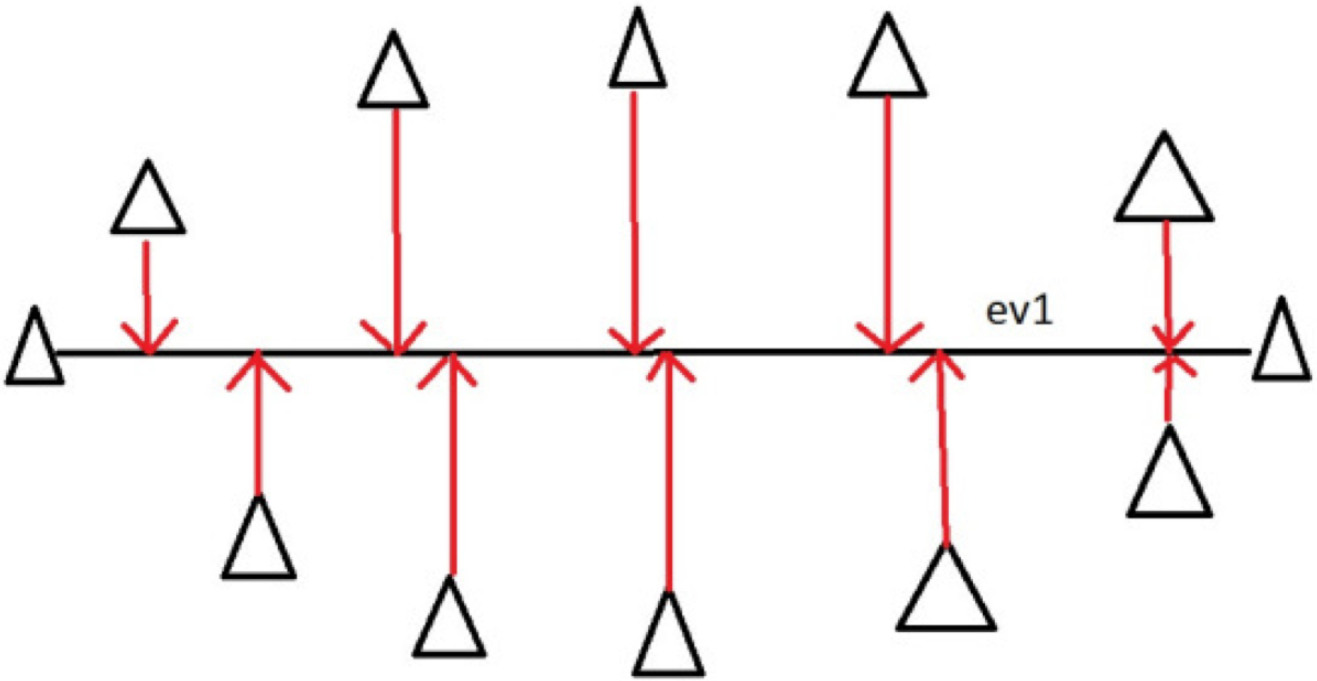
If we choose a vertical line and project the data onto it:



The data isn't very spread out, so it doesn't have a lot of variance. This vertical line is probably not a component.

Example

Horizontal line, data is more spread out. More variance.



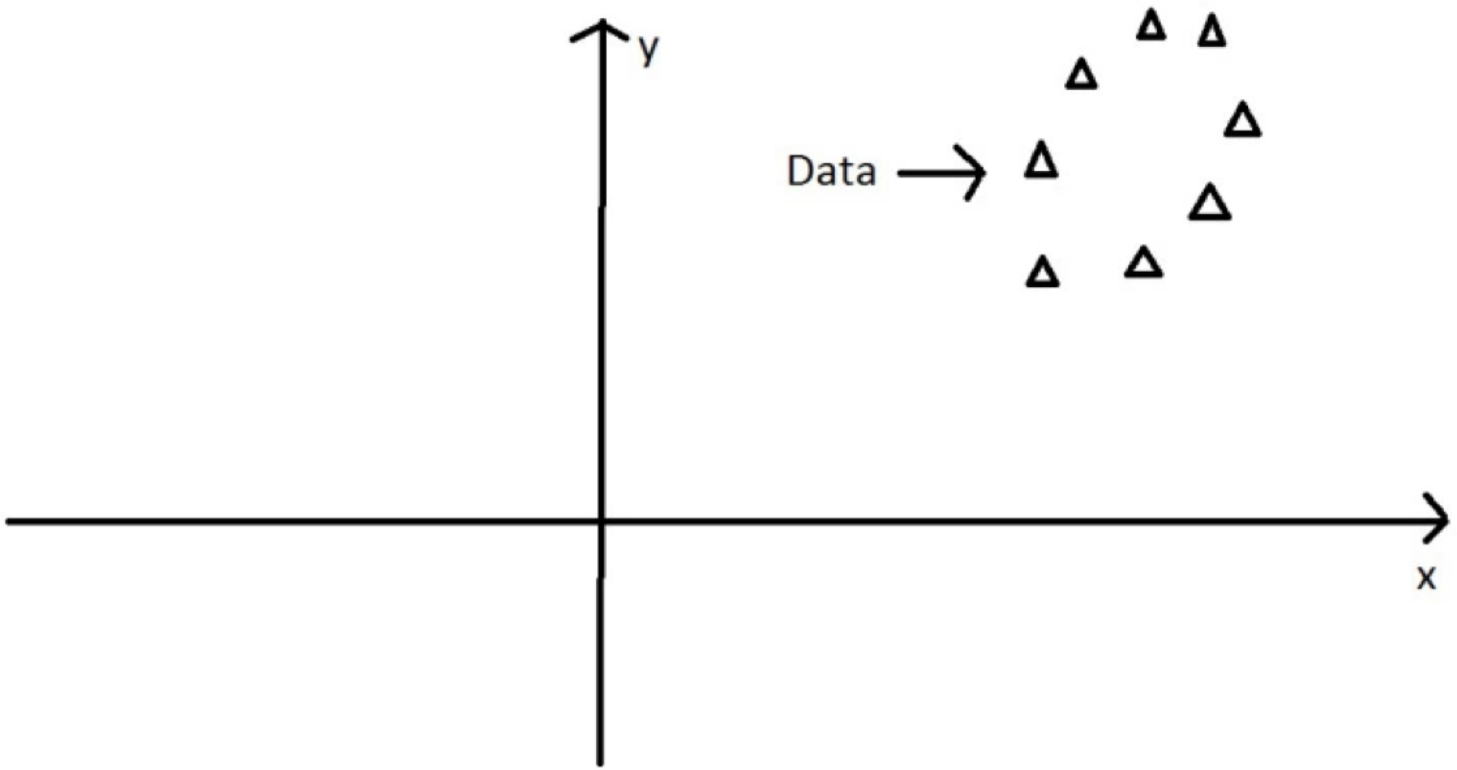
How do we find the principal component? Maths.

Eigenvector

- Eigenvector and Eigenvalues exist as pairs.
- Eigenvector is a direction of a line (vertical, horizontal, 45 degrees etc.).
- Eigenvalue is a number telling you how much variance there is in the data in that direction.
- The Eigenvector with the highest Eigenvalue is the principal component.

Example

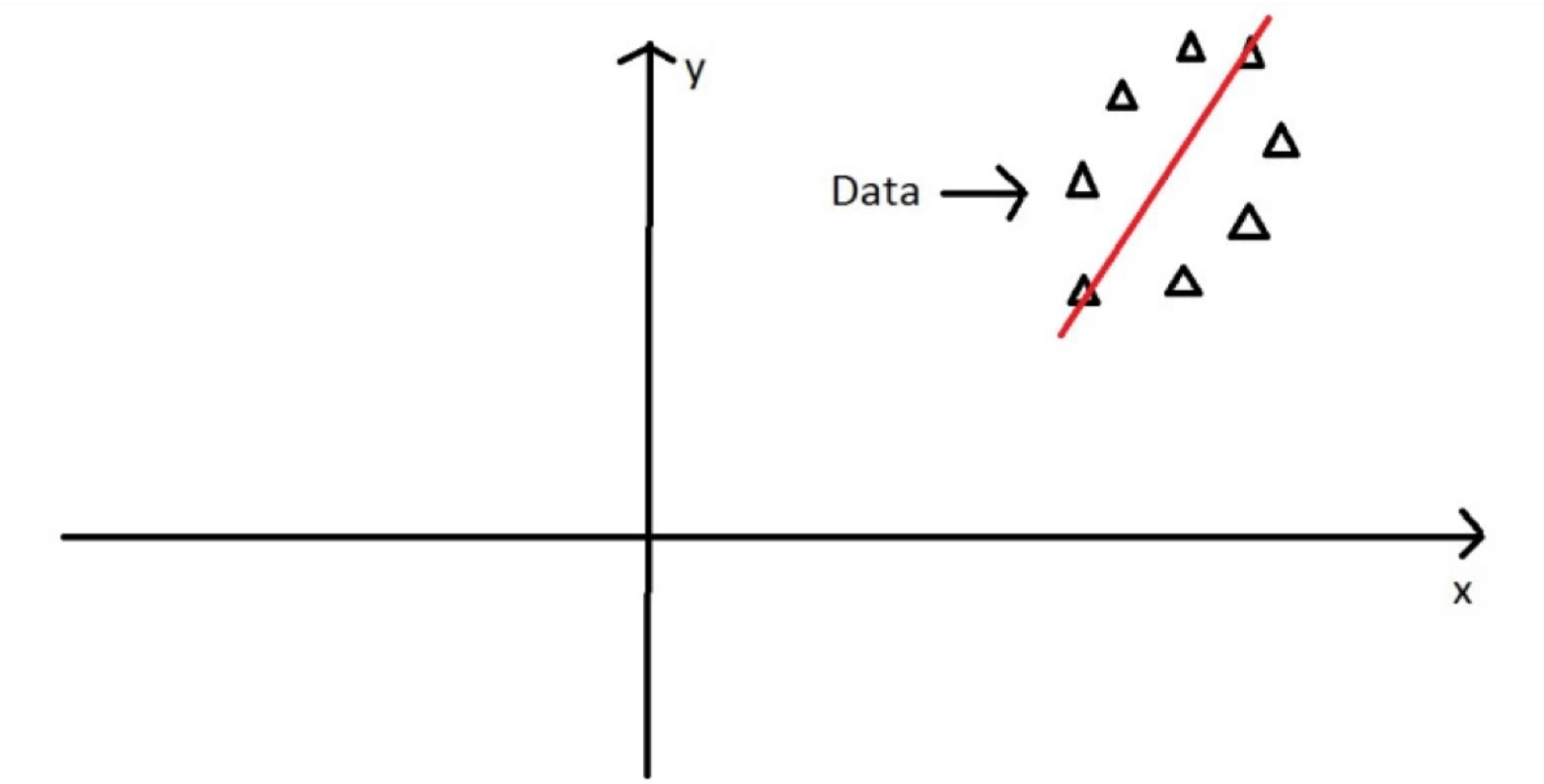
Example: Age and Hrs spent on the internet.
Two variables, two dimensions, two eigenvectors and values.



The Oval here is the data plotted against the x and y axis (x-age, y-hrs on internet).

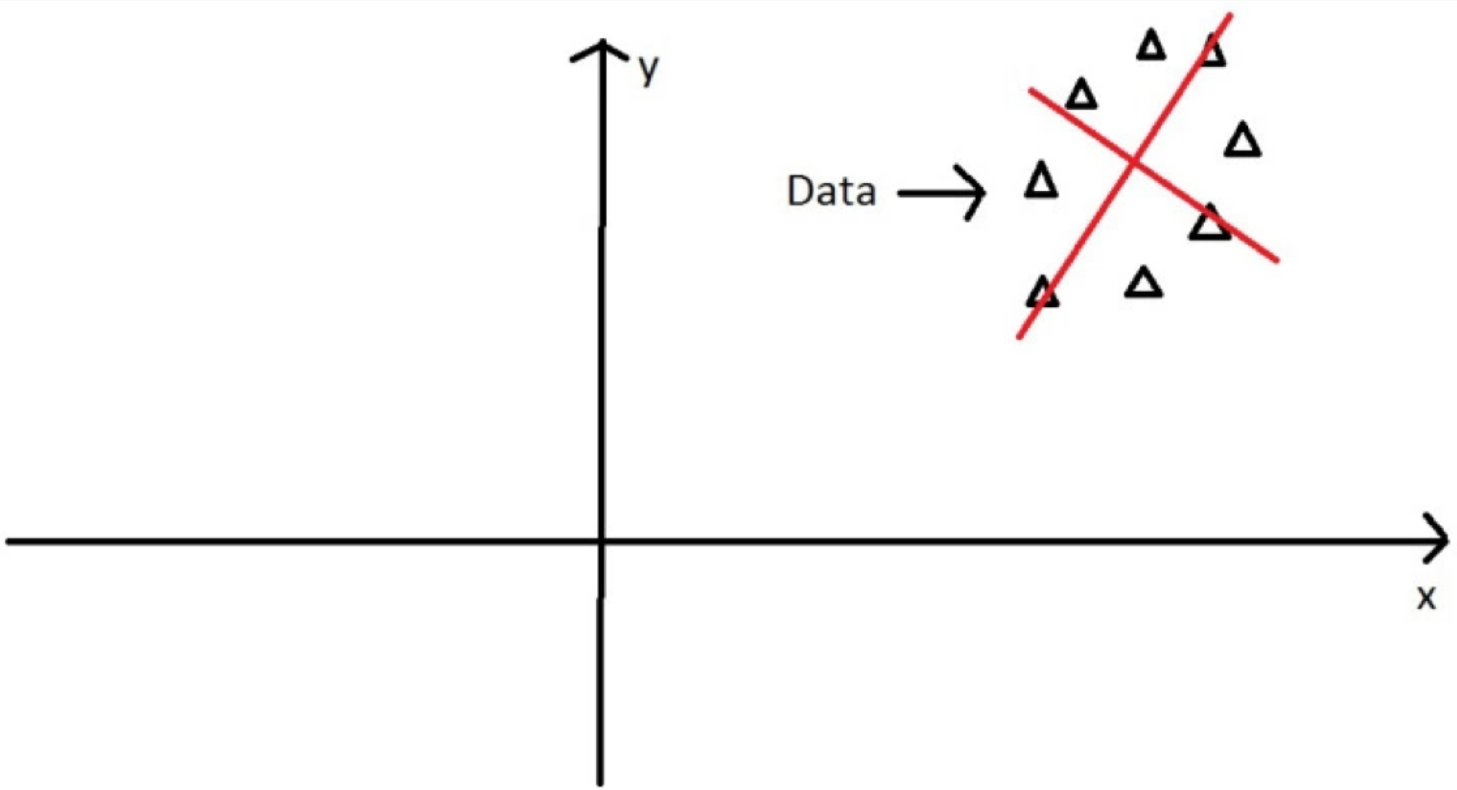
Example

Principal Component is line splitting it lengthways - 1^{st} eigenvector.



Example

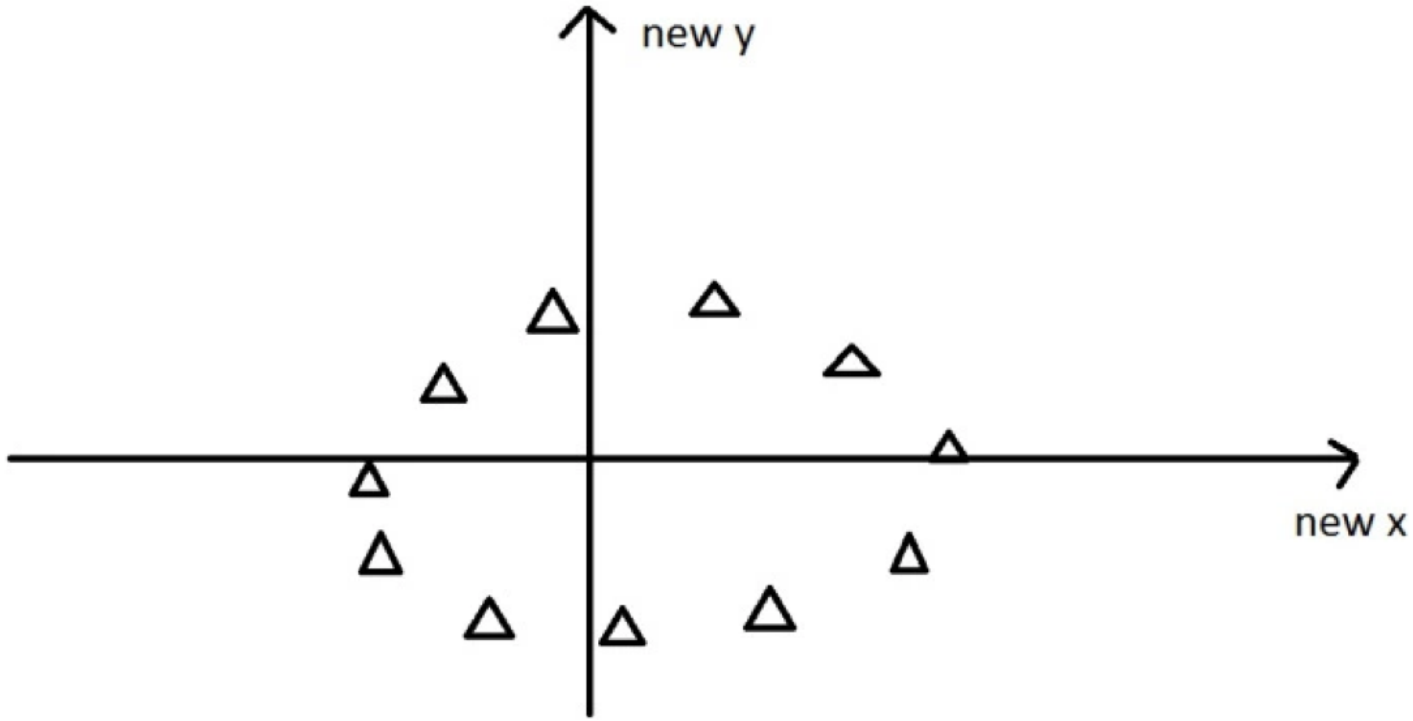
Second is perpendicular to this.



The eigenvectors have to be able to span the whole x-y area, in order to do this (most effectively), the two directions need to be orthogonal (i.e. 90 degrees) to one another. This is why the x and y axis are orthogonal to each other in the first place. It would be really awkward if the y axis was at 45 degrees to the x axis.

Example

If we now project the data against these eigenvectors as axes:



Nothing has been done to the data itself. We're just looking at it from a different angle.

These axes are much more intuitive to the shape of the data now. These directions are where there is most variation, and that is where there is more information.