Feature data is often provided as a *vector*, a quantity that has magnitude and direction. For example, the vectors $\mathbf{a}_1$ and $\mathbf{a}_2$, shown below, each have a length of 3. We've illustrated $\mathbf{a}_1$ with a direction parallel to the y axis and $\mathbf{a}_2$ with a direction $45°$ with respect to the x axis.

$$\mathbf{a}_1 = [2 \quad 6 \quad 1] \qquad \mathbf{a}_2 = [7 \quad 5 \quad 2]$$

Data from several features are often represented as a *matrix*, a tabular representation of a set of numbers as a collection of rows and columns. For example, the matrix $\mathbf{A}$, shown below, is a 2 by 3 (2×3) matrix. The transpose of $\mathbf{A}$ is written as $\mathbf{A}^T$, and is produced by interchanging the rows and columns of $\mathbf{A}$.

$$\mathbf{A} = \begin{bmatrix} 2 & 6 & 1 \\ 7 & 5 & 2 \end{bmatrix} \qquad \mathbf{A}^T = \begin{bmatrix} 2 & 7 \\ 6 & 5 \\ 1 & 2 \end{bmatrix}$$

The eigenvalues and eigenvectors of an m by n matrix are, respectively, the scalar values λ and the vectors $\mathbf{x}$ that are solutions to the following equation:

$$\mathbf{Ax}=\lambda\mathbf{x}.$$

In other words, eigenvectors are the vectors that are unchanged, except for magnitude, when multiplied by $\mathbf{A}$.

The basic equation is $\mathbf{Ax} = \lambda\mathbf{x}$.
The number λ is an eigenvalue of **A**.

Consider matrix $\mathbf{A}$ and vector $\mathbf{x}_1$:

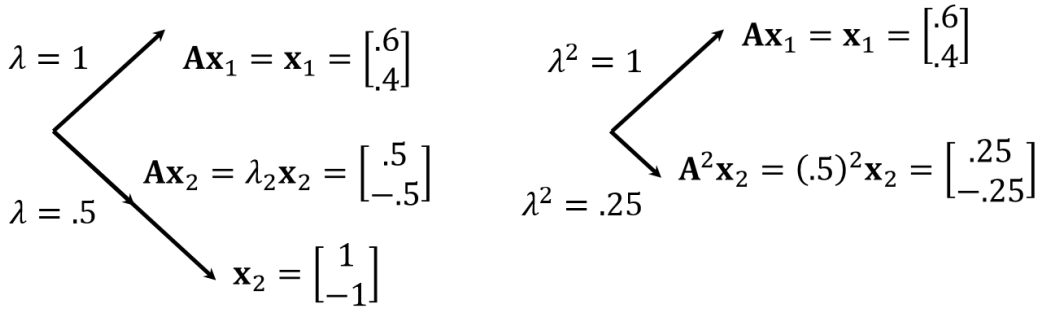
$$\mathbf{A} = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix} \qquad \mathbf{x}_1 = \begin{bmatrix} .6 \\ .4 \end{bmatrix}$$

Notice that if we multiply these two matrices, we obtain $\mathbf{x}_1$ as a result.

$$\mathbf{A}\mathbf{x}_1 = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix} \begin{bmatrix} .6 \\ .4 \end{bmatrix} = \mathbf{x}_1 \qquad (\mathbf{A}\mathbf{x} = \mathbf{x} \text{ means that } \lambda_1 = 1)$$

$$\mathbf{A}\boldsymbol{x}_2 = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} .5 \\ -.5 \end{bmatrix} \qquad (\text{this is } \tfrac{1}{2}\mathbf{x}_2, \text{ so } \lambda_2 = \tfrac{1}{2})$$

Note that if $\mathbf{x}_1$ is multiplied again by $\mathbf{A}$, we still get $\mathbf{x}_1$.



Above, we can see that the eigenvectors keep their directions following this operation, whereas other vectors do not. Also note that all other vectors are combinations of the (two) eigenvectors.

Recall that the variance is the most common measure of the spread of a set of points:

$$Var(x) = s_x^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \bar{x})^2$$

And that the covariance is a measure of the degree to which two variables vary together, and is given by:

$$Cov(x_i, x_j) = \frac{1}{m-1}\sum_{i=1}^{m}(x_{ki} - \overline{x_i})(x_{kj} - \overline{x_j})$$

Where x_{ki} and x_{kj} are the values of the i^{th} and j^{th} feature vectors for the k^{th} object.

The covariance matrix $\mathbf{C}_X$ can be decomposed as follows:

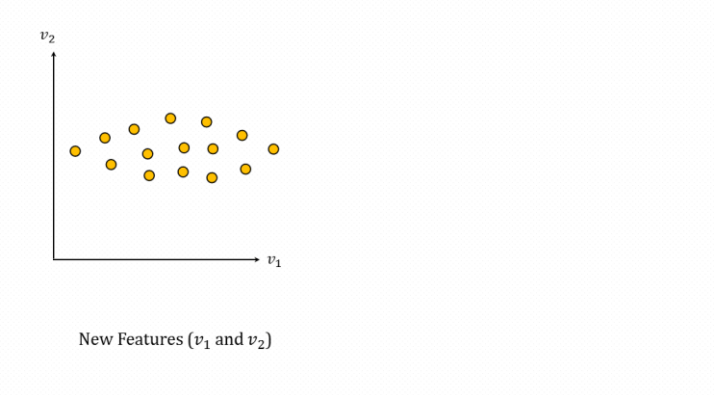
$$\mathbf{C}_X = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^{-1} \quad \mathbf{\Phi}^{-1} = \mathbf{\Phi}^{\mathrm{T}}$$

Thus, $\mathbf{C}_X$ can be decomposed as the product of three matrices. $\mathbf{\Phi}$ is known as the eigenvector matrix and $\mathbf{\Lambda}$ as the eigenvalue matrix. Specifically, *the columns of* $\mathbf{\Phi}$ *are the eigenvectors* of $\mathbf{C}_X$, while *the diagonal elements of* $\mathbf{\Lambda}$ *are the eigenvalues* of $\mathbf{C}_X$.

So what does this all have to do with PCA?

A goal of PCA is to find a transformation of the data that satisfies the following properties:

1. Each pair of new features has 0 covariance.
2. The features are ordered with respect to how much of the variance of the data each new feature captures.
3. The first feature captures as much of the variance of the data as possible.
4. Each successive feature captures as much of the remaining variance as possible, as long as it is orthogonal to the previous components.



PCA: A Mathematical Description

The data matrix D' is the set of transformed data the satisfies the posed goals of PCA.

$$\mathbf{D'} = \mathbf{D}\boldsymbol{\Phi}$$

In words, $\mathbf{D}'$ is the matrix that results from taking the product of the original data matrix, $\mathbf{D}$, with $\boldsymbol{\Phi}$, the eigenvector matrix of $\mathbf{C}$. Recall that $\mathbf{C}$ is the covariance matrix of $\mathbf{D}$ (the product of $\mathbf{D}$ and its transpose, $\mathbf{D}^T$).

The eigenvectors represent linear combinations of the original features, where each is a new axis or dimension. The eigenvalues are measures of the amount of variance captured by the eigenvectors. So that each eigenvector has a corresponding eigenvalue.