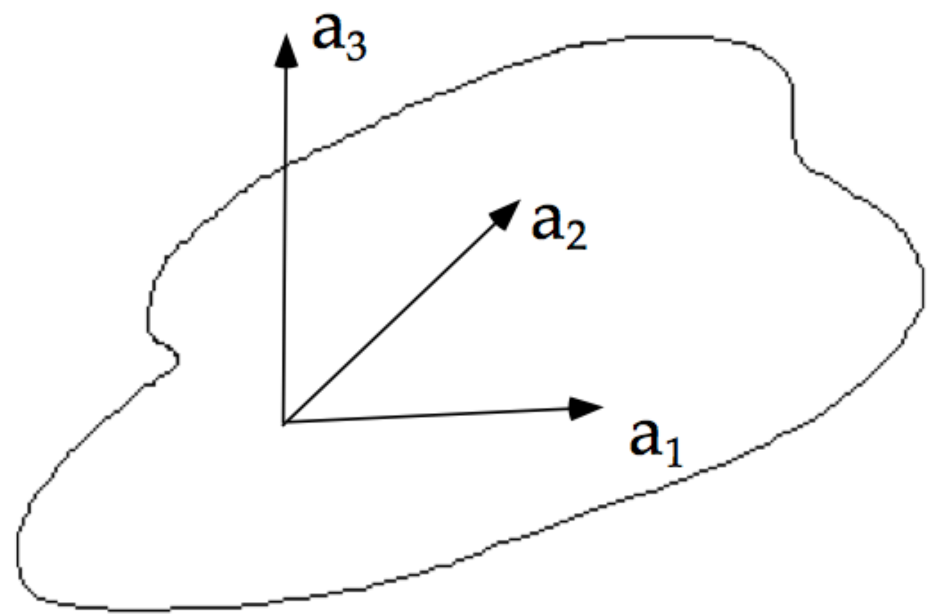
Coordinate Systems, Eigenvectors and Principal Components

Homework: Eigendigits

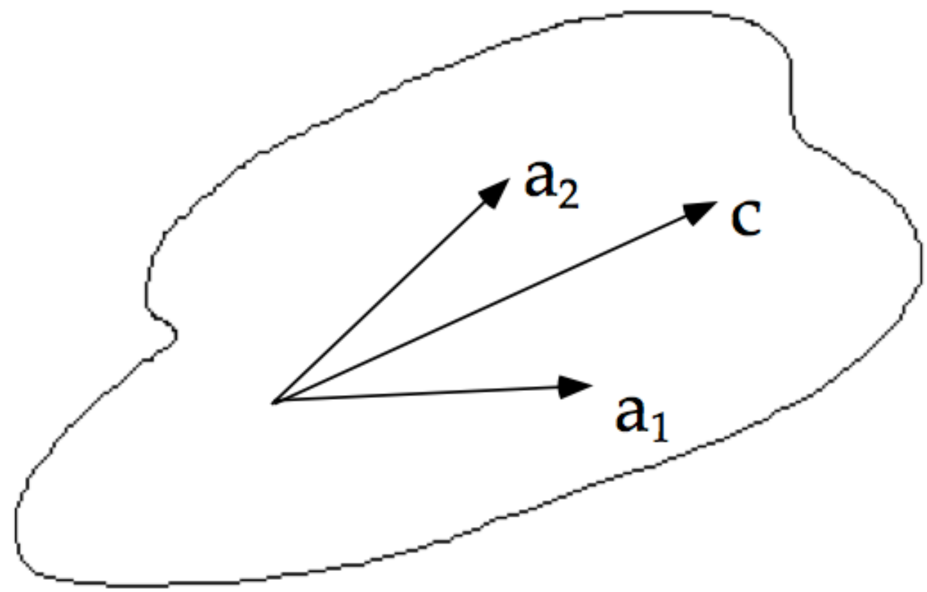
Ax = 0

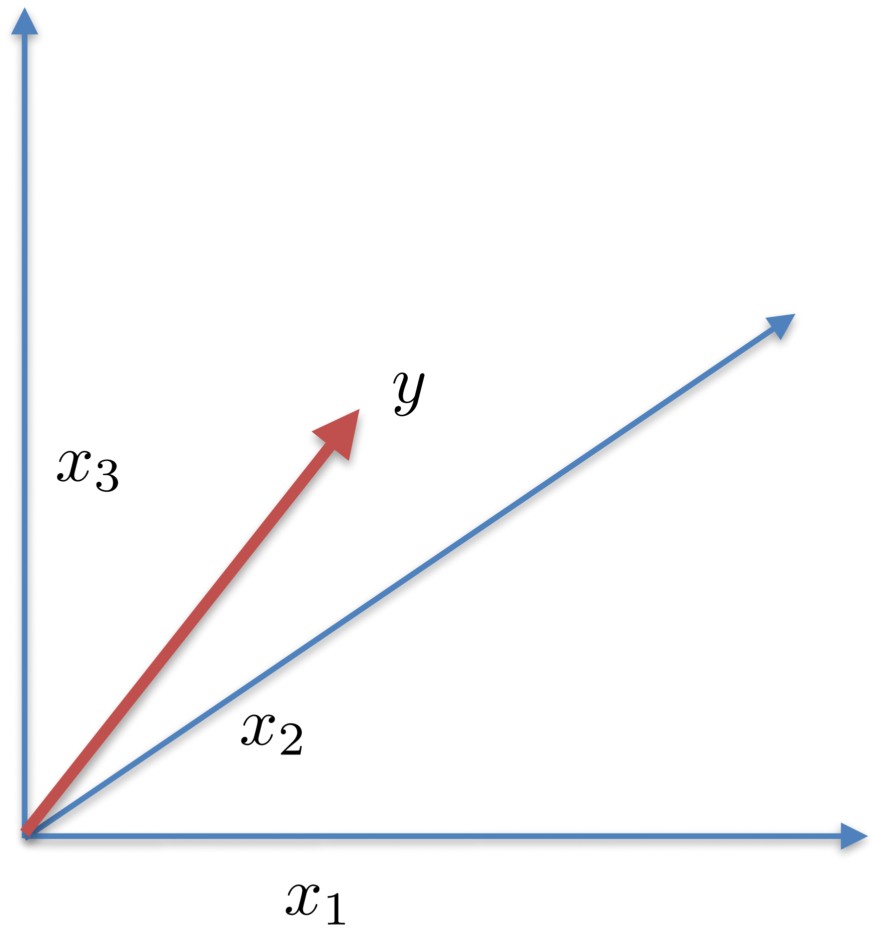
|A| =0



Ax = c

|A| ne 0




y
x_3
x_2
x_1

Can write *y* in terms of the axes vectors

$$y = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} x_1 + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} x_2 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} x_3$$

Can write *y* in terms of the axes vectors

$$y = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} x_1 + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} x_2 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} x_3$$

or equivalently as:

$$y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Here $A = I$ but what about an arbitrary matrix?

Eigenvectors are a special coordinate system

The following example shows a case for the matrix

$$\begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix}$$

where

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

is a special direction for the matrix, since multiplying it by the matrix just results in scaling the vector by a factor $\lambda = 4$; that is,

$$\begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 4 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Solving for eigenvalues' polynomial equation

$$\begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \lambda \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

or, in other words,

$$\begin{bmatrix} 3-\lambda & 1 \\ 2 & 2-\lambda \end{bmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

From the previous section we know that for this equation to have a solution, the columns of the matrix must be linearly dependent, and thus $|W| = 0$. Thus

$$(3-\lambda)(2-\lambda) - 2 = 0$$

which can be solved to find the two eigenvalues $\lambda_1 = 4$ and $\lambda_2 = 1$.

Solving for Eigenvectors

Now for the eigenvectors. Substituting $\lambda_1 = 4$ into the equation results in

$$\begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Now this set of equations is degenerate, meaning that there is only one useful equation in two unknowns. As a consequence there is an infinity of solutions, and you must pick one arbitrarily. Pick $v_1 = 1$. Then $v_2 = 1$. Thus the eigenvector associated with $\lambda_1 = 4$ is

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Suppose that the coordinate transformation is given by

$$\boldsymbol{x}^* = A\boldsymbol{x}$$

$$\boldsymbol{y}^* = A\boldsymbol{y}$$

Given the transformation

$$\boldsymbol{y} = W\boldsymbol{x}$$

what happens to W when the coordinate system is changed to the starred system? That is, for some W^* it will be true that

$$\boldsymbol{y}^* = W^*\boldsymbol{x}^*$$

Similar Transformations

What is the relation between W and W^*? One way to find out is to change back to the original system, transform by W, and then transform back to the starred system; that is,

$$\boldsymbol{x} = A^{-1}\boldsymbol{x}^*$$

$$\boldsymbol{y} = W\boldsymbol{x}$$

$$\boldsymbol{y}^* = A\boldsymbol{y}$$

Putting these transformations together:

$$\boldsymbol{y}^* = AWA^{-1}\boldsymbol{x}^*$$

Since the vector transformation taken by the two different routes should be the same, it must be true that

$$W^* = AWA^{-1}$$

Matrices related in this way are called *similar*.

Now let's relate this discussion to eigenvectors. Suppose that the eigenvectors have been chosen as the basis set. Then for a given eigenvector $\boldsymbol{y}_i$,

$$W\boldsymbol{y}_i = \lambda\boldsymbol{y}_i$$

and if Y is a matrix whose columns are the eigenvectors $\boldsymbol{y}_i$, then

$$WY = Y\Lambda$$

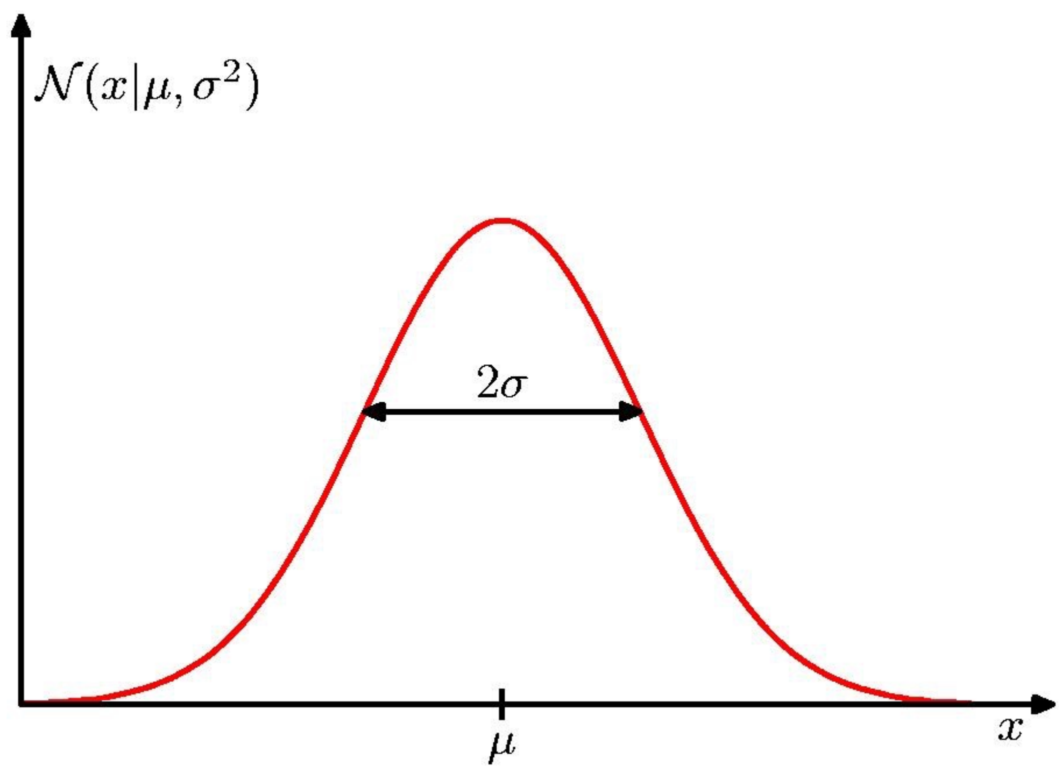
Here Λ is a matrix whose only nonzero components are the diagonal elements λ_i. Premultiplying both sides by Y^{-1},

$$Y^{-1}WY = \Lambda$$

What this equation means is that given a matrix W, the transformation it defines can always be simplified to that of a matrix whose only nonzero elements are diagonal by transforming to coordinates that use its eigenvectors as a basis. Furthermore, those elements are the eigenvalues.

The Gaussian Distribution

$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$



$$\mathcal{N}(x|\mu,\sigma^2) > 0$$

$$\int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right)\,\mathrm{d}x = 1$$

Gaussian Mean and Variance

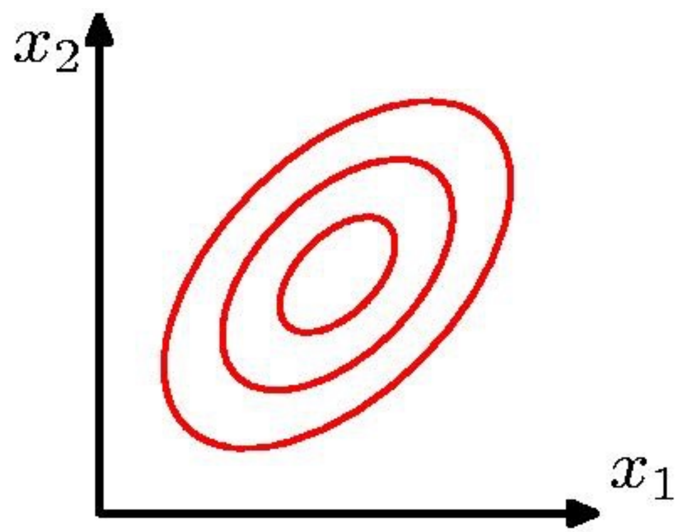
$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) x \,\mathrm{d}x = \mu$$

$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) x^2 \,\mathrm{d}x = \mu^2 + \sigma^2$$

$$\mathrm{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \sigma^2$$

The Multivariate Gaussian

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{ -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu}) \right\}$$



Mean and Variance in the vector case

The mean vector is defined by

$$\boldsymbol{M} = E\{\boldsymbol{X}\} = \int \boldsymbol{X} p(\boldsymbol{X}) d\boldsymbol{X}$$

and the covariance matrix by

$$\Sigma = E\{(\boldsymbol{X} - M)(\boldsymbol{X} - M)^T\}$$

In practice, with real data you will use the sample mean vector and sample covariance matrix. Where $\boldsymbol{X}^k, k = 1, N$ are the samples,

$$M = \frac{1}{N} \sum_{k=1}^{N} X^k$$

$$\Sigma = \frac{1}{N} \sum_{k=1}^{N} (\boldsymbol{X}^k - M)(\boldsymbol{X}^k - M)^T$$

Example

Suppose

$$\boldsymbol{X}^1 = \begin{pmatrix} -1 \\ 3 \\ 1 \end{pmatrix},\ \boldsymbol{X}^2 = \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix},\ \boldsymbol{X}^3 = \begin{pmatrix} 2 \\ 2 \\ 3 \end{pmatrix}$$

Then the mean value is

$$M = \frac{1}{3}\begin{pmatrix} 3 \\ 6 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

So that

$$\boldsymbol{X}^1 - M = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \boldsymbol{X}^2 - M = \begin{pmatrix} 1 \\ -1 \\ -2 \end{pmatrix}, \boldsymbol{X}^3 - M = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$$

and the covariance matrix is given by

$$\Sigma = \frac{1}{3}\left\{\begin{bmatrix} 4 & -2 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & -1 & -2 \\ -1 & 1 & 2 \\ -2 & 2 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 2 \\ 0 & 0 & 0 \\ 2 & 0 & 4 \end{bmatrix}\right\}$$

$$= \frac{1}{3}\begin{bmatrix} 6 & -3 & 0 \\ -3 & 2 & 2 \\ -2 & 2 & 8 \end{bmatrix}$$

$$\boldsymbol{\Sigma}\mathbf{u}_i = \lambda_i \mathbf{u}_i \tag{2.45}$$

where $i = 1, \ldots, D$. Because $\boldsymbol{\Sigma}$ is a real, symmetric matrix its eigenvalues will be real, and its eigenvectors can be chosen to form an orthonormal set, so that

$$\mathbf{u}_i^{\mathrm{T}} \mathbf{u}_j = I_{ij} \tag{2.46}$$

where I_{ij} is the i, j element of the identity matrix and satisfies

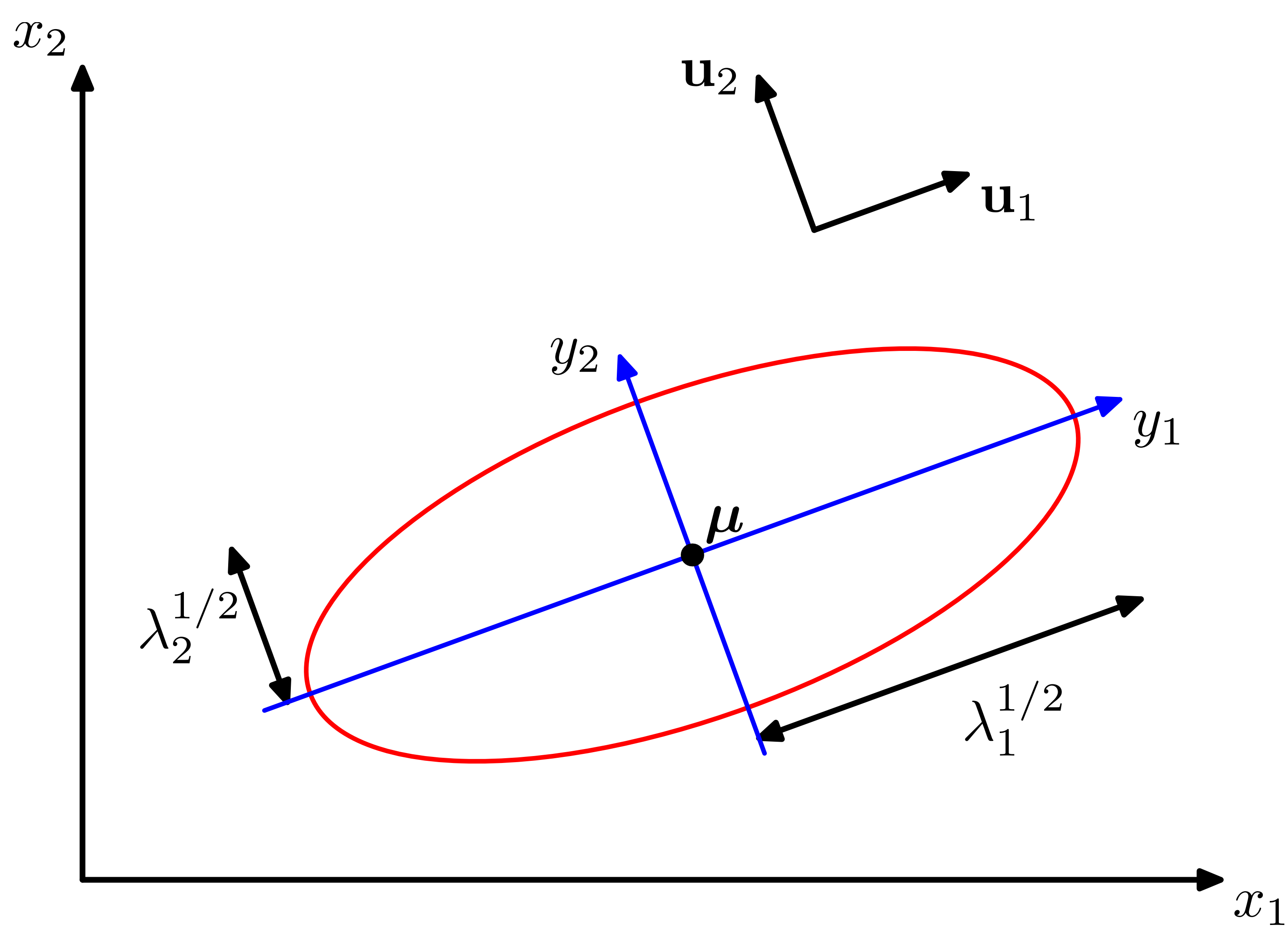
$$I_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise.} \end{cases} \tag{2.47}$$

The covariance matrix $\boldsymbol{\Sigma}$ can be expressed as an expansion in terms of its eigenvectors in the form

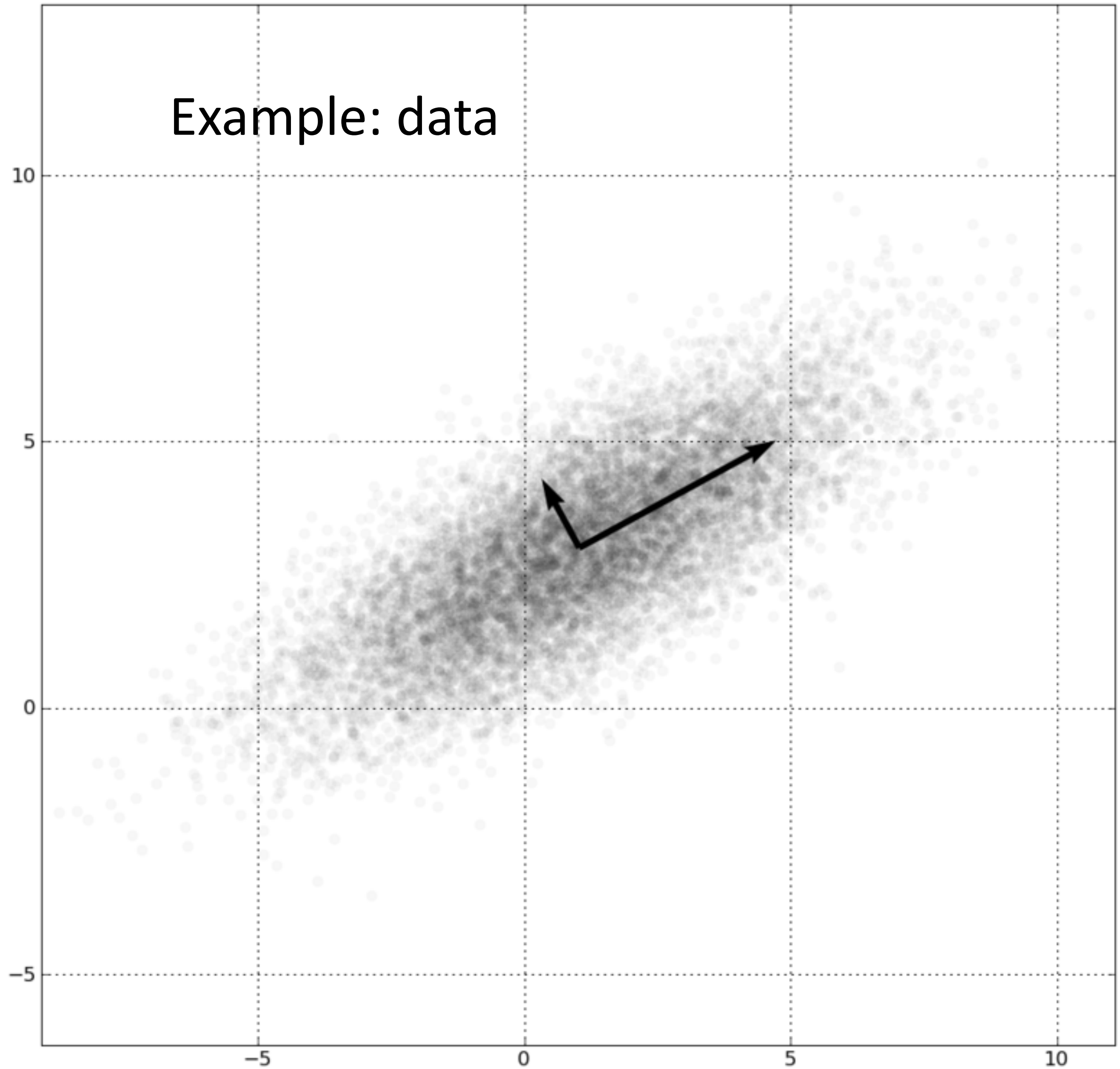
$$\boldsymbol{\Sigma} = \sum_{i=1}^{D} \lambda_i \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}} \tag{2.48}$$

and similarly the inverse covariance matrix $\boldsymbol{\Sigma}^{-1}$ can be expressed as

$$\boldsymbol{\Sigma}^{-1} = \sum_{i=1}^{D} \frac{1}{\lambda_i} \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}}. \tag{2.49}$$


x_2
$\mathbf{u}_2$
$\mathbf{u}_1$
y_2
y_1
$\boldsymbol{\mu}$
$\lambda_2^{1/2}$
$\lambda_1^{1/2}$
x_1

Example: data



High-Dimensional Spaces

All of these operations will be made clear with an example after one more development. Suppose, as is often the case, that the dimension of the space is extremely large. Now the standard way to proceed would be to choose eigenvectors u_k and eigenvalues λ_k of the sample covariance matrix Σ where

$$\begin{aligned}\Sigma &= \frac{1}{M}\sum_{n=1}^{M} \boldsymbol{X}_n \boldsymbol{X}_n^T \\ &= AA^T\end{aligned}$$

(After subtracting mean)

where

$$A = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_M]$$

an $M \times N$ matrix of M data samples. The problem with this tack is that it is infeasible owing to the high dimensionality of the matrix Σ. Since for an image the dimension of $\boldsymbol{X}$ is n^2, then the dimension of Σ is $n^2 \times n^2$. For typical values of n, say 256, this is impossibly large. Salvation comes from the fact that the matrix Σ may be approximated by a matrix of lower rank. That is, most of the variation can be captured by projecting the data onto a subspace whose dimension is much less than the dimension of the space.

Rather than finding the eigenvectors of the larger system, consider finding the eigenvectors of the $M \times M$ system

$$A^T A \boldsymbol{v} = \mu \boldsymbol{v} \tag{6}$$

Rather than finding the eigenvectors of the larger system, consider finding the eigenvectors of the $M \times M$ system

$$A^T A\boldsymbol{v} = \mu\boldsymbol{v} \tag{6}$$

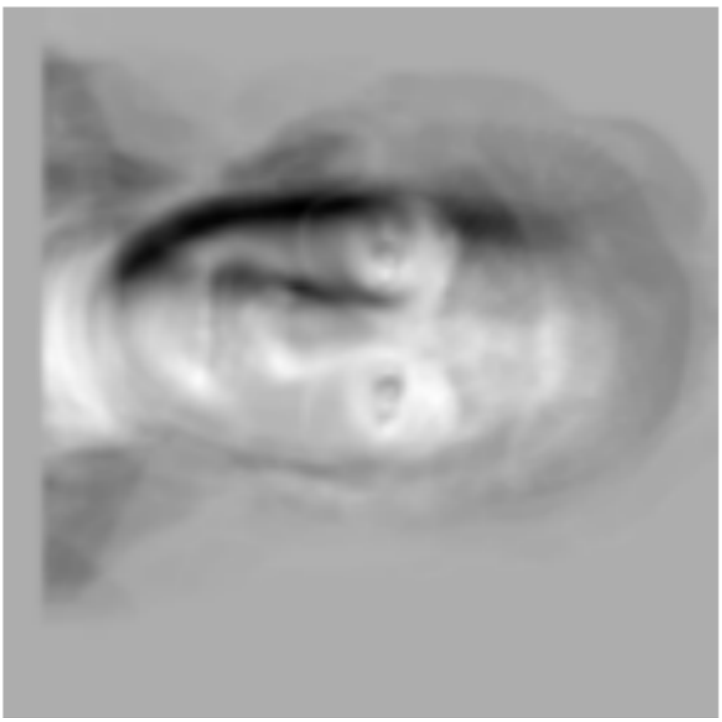
Premultiplying both sides by A,

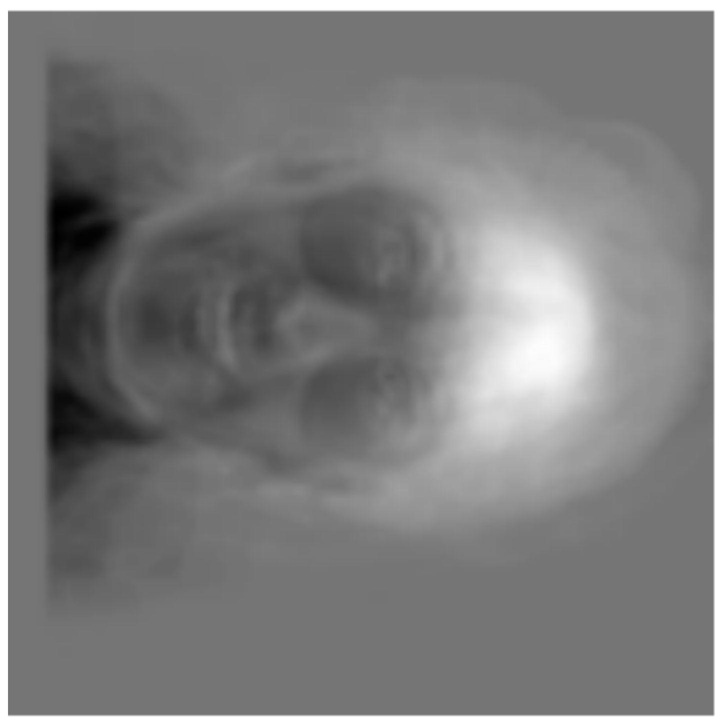
$$AA^T A\boldsymbol{v} = \mu A\boldsymbol{v}$$

What this equation shows is that if $\boldsymbol{v}$ is an eigenvector of $A^T A$, then $A\boldsymbol{v}$ is an eigenvector of Σ. Furthermore, the eigenvalues of the smaller system are the same as those of the much larger system. It turns out also that these are the M largest eigenvalues. So to find the eigenvectors of the larger system, first find the eigenvalues and eigenvectors of the smaller system, and then multiply the eigenvectors by A.

The face image is described by an $N \times N$ array of brightness values. Given M exemplars of images with known identities, the objective is to take a new image and identify it with the training image to which it is most similar. The key element is in the similarity metric. It should be chosen to score the essential variations in the data. From the last section, the way to discover the essential variations is with principal components analysis, which identifies the eigenvalues of the covariance matrix of all the data.





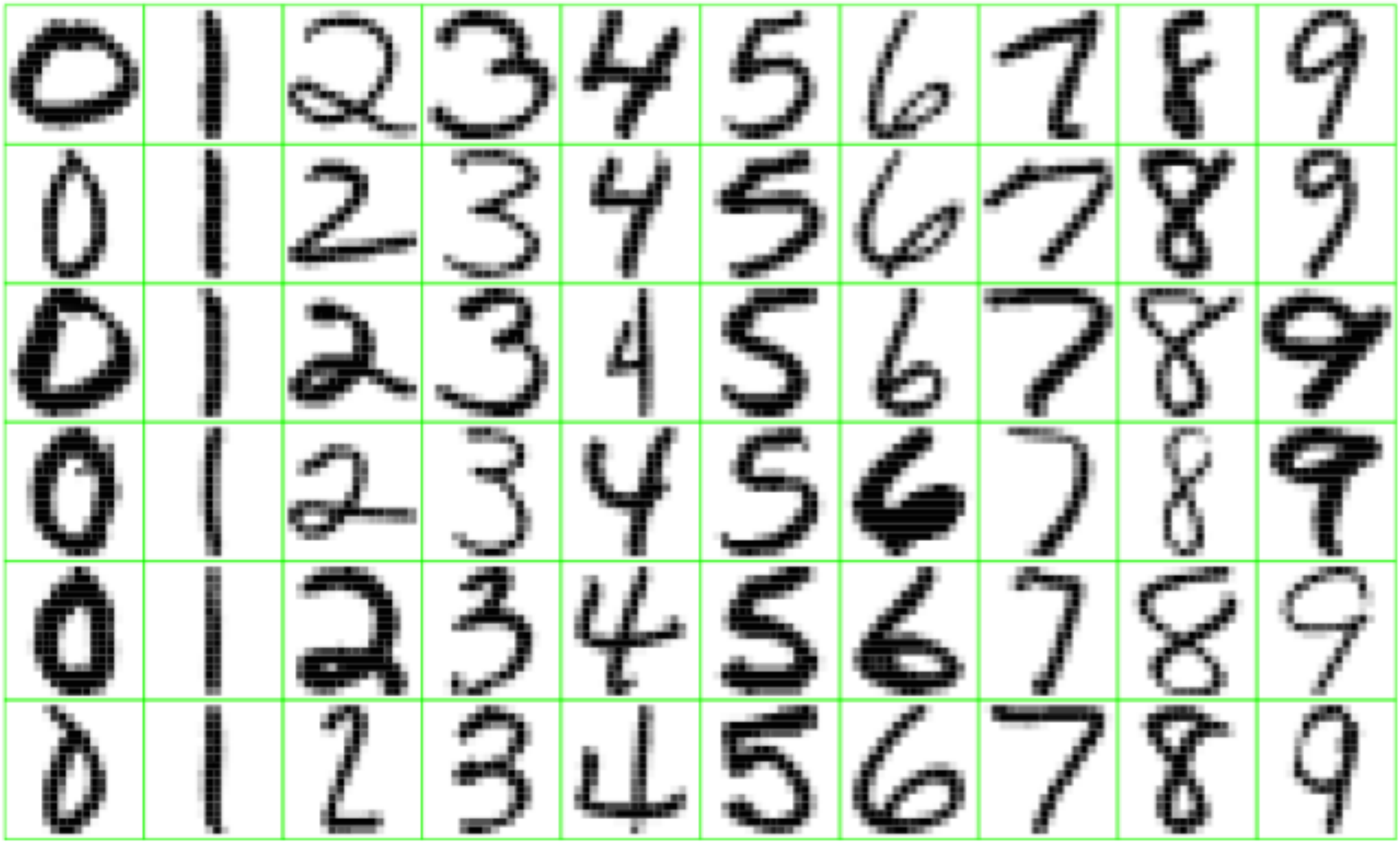


FIGURE 1.2. *Examples of handwritten digits from U.S. postal envelopes.*

To begin, identify the training set of images as $\boldsymbol{I}_1, \boldsymbol{I}_2, \ldots, \boldsymbol{I}_M$. To work with these it is useful to subtract the bias introduced, as all the brightness levels are positive. Thus first identify the "average face"

$$\boldsymbol{I}_{ave} = \frac{1}{M}\sum_{n=1}^{M} \boldsymbol{I}_n$$

and then convert the training set by subtracting the average,

$$\boldsymbol{X}_i = \boldsymbol{I}_i - \boldsymbol{I}_{ave}, i = 1, \ldots, M$$

Now use Equation 6 to find M eigenvectors v_k and eigenvalues λ_k.

From the "short" eigenvectors (of length M), the larger eigenvectors $\boldsymbol{u}_k$, termed eigenfaces, can be constructed using $\boldsymbol{v}_k$, as follows:

$$\boldsymbol{u}_i = \sum_{k=1}^{M} v_{ik}\boldsymbol{X}_k$$

Now that the direction principal variations have been calculated in the space of faces, this information can be used to classify a new face in terms of the faces in the data set. To do so, compute the coordinates of the new image in the eigenvector space $\boldsymbol{\Omega} = (\omega_1, \omega_2, \cdots, \omega_M)$ as follows:

$$\omega_k = u_k^T(\boldsymbol{I} - \boldsymbol{I}_{ave}), k = 1, \cdots, M$$

Next compare $\boldsymbol{\Omega}$ to the $\boldsymbol{\Omega}$s for each of the classes to pick the closest; that is, pick the class k that minimizes

$$||\boldsymbol{\Omega} - \boldsymbol{\Omega}_k||$$

Homework

By computing m eigenvectors, you can get n^2 dimensional vectors

Test a classifier that uses k nearest neighbors

Report on the best settings for m and k
for a given number of training and test samples

Generate representative eigendigits