ECE 417 Lecture 5: Principal Component Analysis (PCA)

Mark Hasegawa-Johnson

9/6/2019

Content

- Linear transforms
- Eigenvectors
- Eigenvalues
- Symmetric matrices
- Gaussian random vectors
- Principal component axes = eigenvectors of the covariance
- Gram matrix
- Singular value decomposition

Linear Transforms

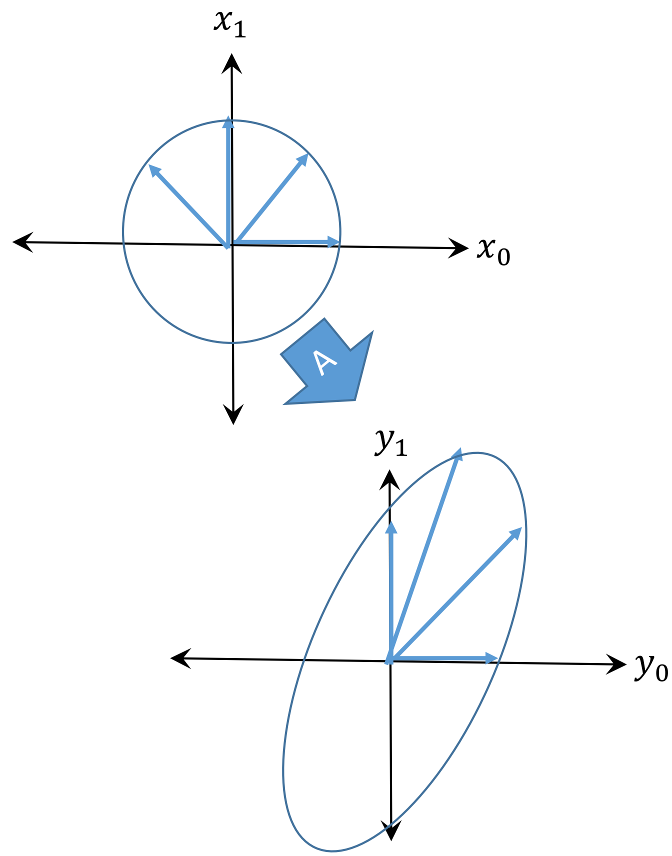
A linear transform $\vec{y} = A\vec{x}$ maps vector space $\vec{x}$ onto vector space $\vec{y}$.

For example: the matrix $A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$ maps the vectors

$$\vec{x}_0, \vec{x}_1, \vec{x}_2, \vec{x}_3 \ = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

to the vectors

$$\vec{y}_0, \vec{y}_1, \vec{y}_2, \vec{y}_3 \ = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ \sqrt{2} \end{bmatrix}$$



Linear Transforms

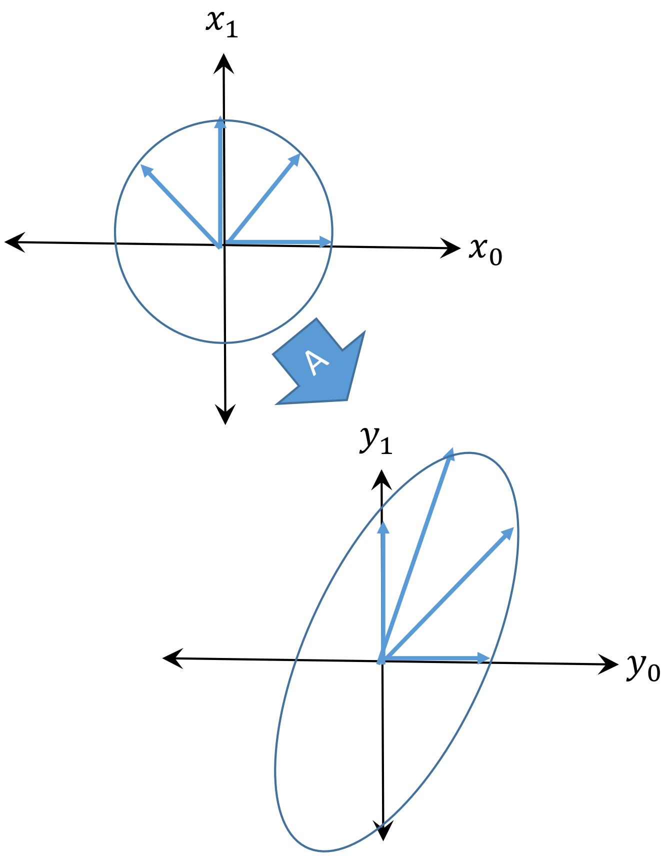
A linear transform $\vec{y} = A\vec{x}$ maps vector space $\vec{x}$ onto vector space $\vec{y}$.

For example: the matrix $A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$ maps the vectors

$$X = \begin{bmatrix} 1 & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 1 & \frac{1}{\sqrt{2}} \end{bmatrix}$$

to the vectors

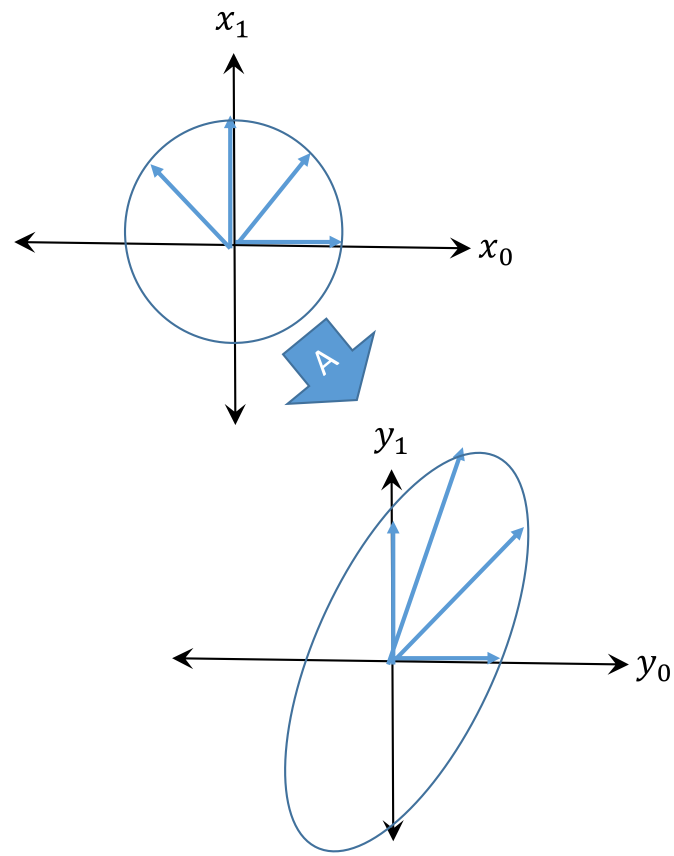
$$Y = \begin{bmatrix} 1 & \sqrt{2} & 1 & 0 \\ 0 & \sqrt{2} & 2 & \sqrt{2} \end{bmatrix}$$



Linear Transforms

A linear transform $\vec{y} = A\vec{x}$ maps vector space $\vec{x}$ onto vector space $\vec{y}$. The absolute value of the determinant of A tells you how much the area of a unit circle is changed under the transformation.

For example: if $A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$, then the unit circle in $\vec{x}$ (which has an area of π) is mapped to an ellipse with an area of $\pi\ abs(|A|) = 2\pi$.



Content

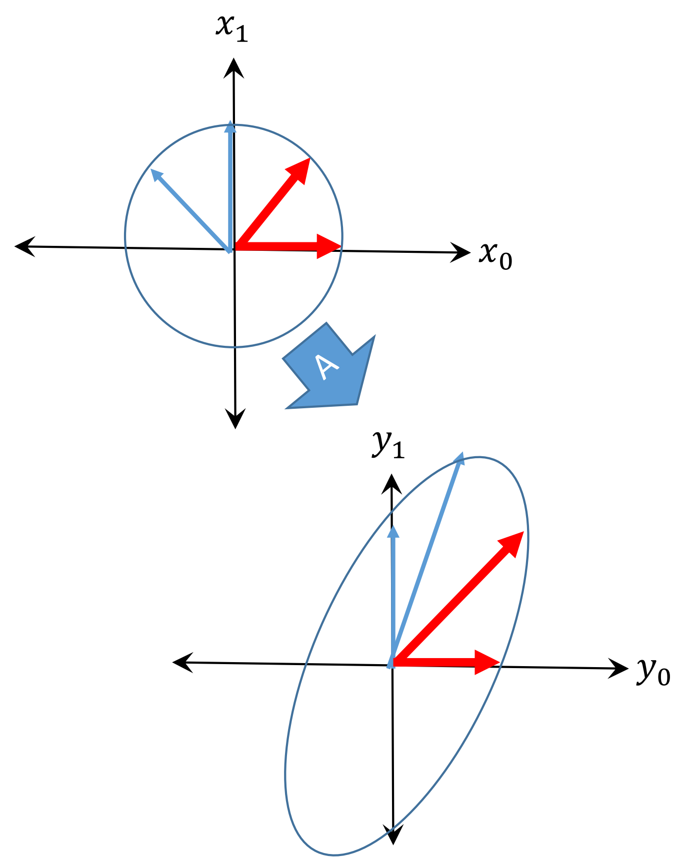
- Linear transforms
- Eigenvectors
- Eigenvalues
- Symmetric matrices
- Gaussian random vectors
- Principal component axes = eigenvectors of the covariance
- Gram matrix
- Singular value decomposition

Eigenvectors

- For a D-dimensional square matrix, there may be up to D different directions $\vec{x} = \vec{v}_d$ such that, for some scalar λ_d,
$$A\vec{v}_d = \lambda_d \vec{v}_d$$
- For example: if $A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$, then the eigenvectors and eigenvalues are

$$\vec{v}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \vec{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \lambda_0 = 1, \lambda_1 = 2$$

- Those vectors are red and extra-thick, in the figure to the left. Notice that one of the vectors gets scaled by $\lambda_0 = 1$, but the other gets scaled by $\lambda_1 = 2$.

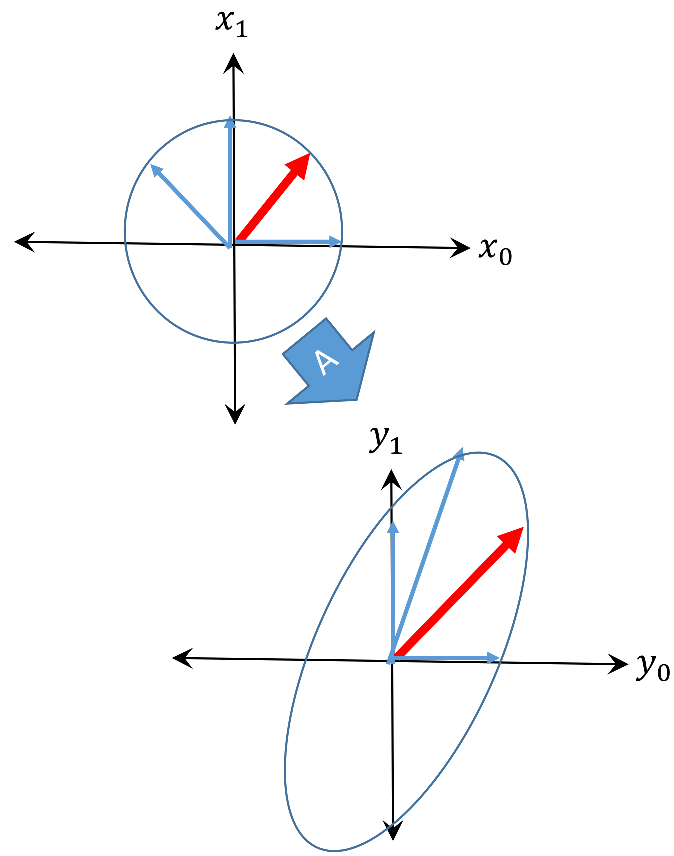


Eigenvectors

- An eigenvector is a direction, not just a vector. That means that if you multiply an eigenvector by any scalar, you get the same eigenvector: if $A\vec{v}_d = \lambda_d \vec{v}_d$, then it's also true that $cA\vec{v}_d = c\lambda_d \vec{v}_d$ for any scalar c.
- For example: the following are all the same eigenvector

$$\vec{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \sqrt{2}\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, -\vec{v}_1 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

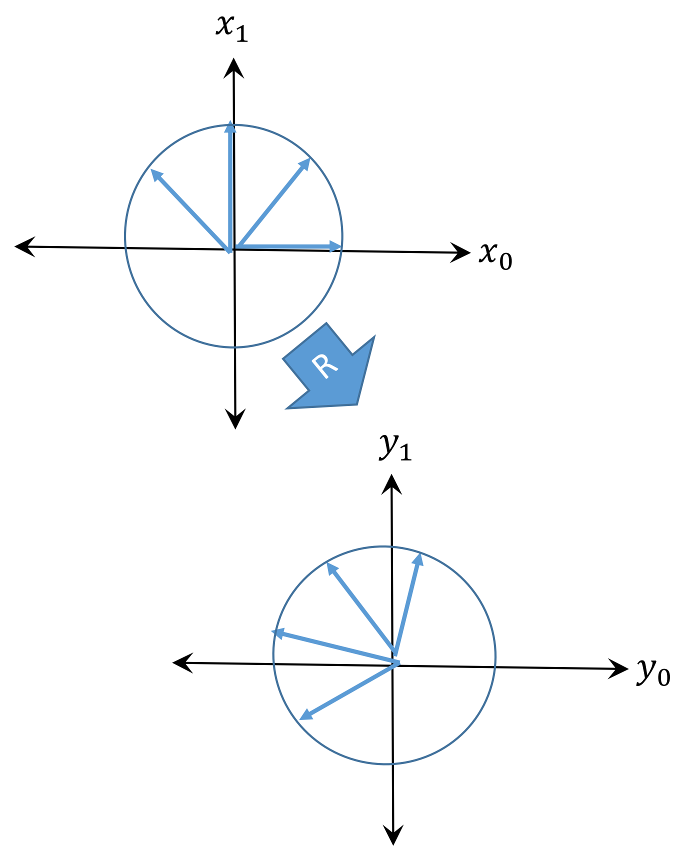
- Since **<u>scale and sign don't matter</u>**, by convention, we normalize so that
  - An eigenvector is always unit-length: $\|\vec{v}_d\|_2 = 1$ and
  - the first nonzero element is non-negative: $v_d[0] \geq 0$



Eigenvectors

- Notice that only square matrices can have eigenvectors. For a non-square matrix, the equation $A\vec{v}_d = \lambda_d \vec{v}_d$ is impossible --- the dimension of the output is different from the dimension of the input.
- Not all matrices have eigenvectors! For example, a rotation matrix doesn't have any real-valued eigenvectors:

$$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

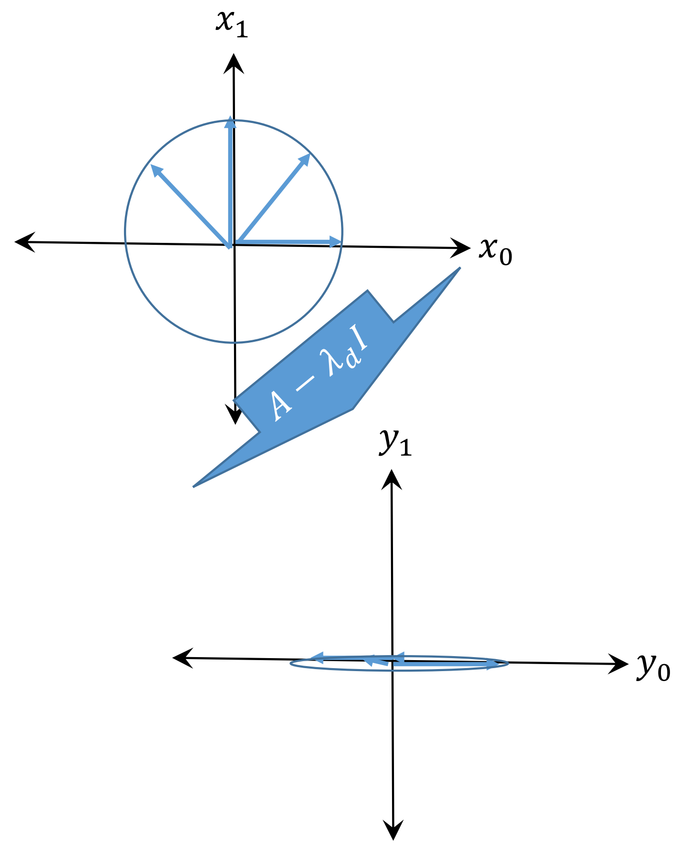


Content

- Linear transforms
- Eigenvectors
- Eigenvalues
- Symmetric matrices
- Gaussian random vectors
- Principal component axes = eigenvectors of the covariance
- Gram matrix
- Singular value decomposition

Eigenvalues

$$A\vec{v}_d = \lambda_d \vec{v}_d$$
$$A\vec{v}_d = \lambda_d I \vec{v}_d$$
$$A\vec{v}_d - \lambda_d I \vec{v}_d = \vec{0}$$
$$(A - \lambda_d I)\vec{v}_d = \vec{0}$$



That means that when you use the linear transform $(A - \lambda_d I)$ to transform the unit circle, the result has an area of:

$$\pi|A - \lambda I| = 0$$

Example:

$$|A - \lambda I| = \begin{vmatrix} 1-\lambda & 1 \\ 0 & 2-\lambda \end{vmatrix} = 2 - 3\lambda + \lambda^2$$

…which has roots at $\lambda_0 = 1, \lambda_1 = 2$.

Eigenvalues

Let's talk about that equation, $|A - \lambda I| = 0$. Remember how the determinant is calculated, for example if

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}, \text{ then } |A - \lambda I| = 0 \text{ means that}$$

$$0 = |A - \lambda I| = \begin{vmatrix} a-\lambda & b & c \\ d & e-\lambda & f \\ g & h & i-\lambda \end{vmatrix} =$$

$$(a-\lambda)(e-\lambda)\,(i-\lambda) - b(d(i-\lambda) - gf) + c\big(dh - g(e-\lambda)\big)$$

- We assume that $a, b, c, d, e, f, g, h, i$ are all given in the problem statement. Only λ is unknown. So the equation $|A - \lambda I| = 0$ is a D'th order polynomial in one variable.
- The fundamental theorem of algebra says that a D'th order polynomial ALWAYS has D roots (counting repeated roots and complex roots).

Eigenvalues

So a DxD matrix always has D eigenvalues (counting complex and repeated eigenvalues). This is true even if the matrix has no eigenvectors!! The eigenvalues are the D solutions of the polynomial equation

$$|A - \lambda I| = 0$$

Summary:

- Not every square matrix has eigenvectors, but…
- Every DxD square matrix has exactly D eigenvalues (counting possibly complex eigenvalues, and repeated eigenvalues).

Content

- Linear transforms
- Eigenvectors
- Eigenvalues
- Symmetric matrices
- Gaussian random vectors
- Principal component axes = eigenvectors of the covariance
- Gram matrix
- Singular value decomposition

Left and right eigenvectors

We've been working with right eigenvectors and right eigenvalues:

$$A\vec{v}_d = \lambda_d \vec{v}_d$$

There may also be left eigenvectors, which are row vectors $\vec{u}_d^T$, and corresponding left eigenvalues μ_d:

$$\vec{u}_d^T A = \kappa_d \vec{u}_d^T$$

Eigenvectors on both sides of the matrix

You can do an interesting thing if you multiply the matrix by its eigenvectors both before and after:

$$\vec{u}_i^T A\vec{v}_j = \vec{u}_i^T\left(\lambda_j \vec{v}_j\right) = \lambda_j \vec{u}_i^T \vec{v}_j$$

… but …

$$\vec{u}_i^T A\vec{v}_j = \left(\kappa_i \vec{u}_i^T\right)\vec{v}_j = \mu_i \vec{u}_i^T \vec{v}_j$$

There are only two ways that both of these things can be true. Either

$$\kappa_i = \lambda_j$$

… or …

$$\vec{u}_i^T \vec{v}_j = 0$$

Left and right eigenvectors must be paired!!

Remember that eigenvalues are the D solutions of the polynomial equation $|A - \lambda_d I| = 0$. Almost always, these D solutions are all different! In that case, the left and right eigenvectors must be paired so that

$$\vec{u}_i^T \vec{v}_j = 0 \quad i \neq j$$

and

$$\kappa_i = \lambda_i = i^{th}\ solution\ of\ \ |A - \lambda I| = 0$$

Symmetric matrices: left == right

If A is symmetric ($A = A^T$), then the left and right eigenvectors and eigenvalues are the same, because

$$\lambda_d \vec{u}_d^T = \vec{u}_d^T A = (A^T \vec{u}_d)^T = (A\vec{u}_d)^T$$

… and that last term is only equal to $\lambda_d \vec{u}_d^T$ IF AND ONLY IF $\vec{u}_d = \vec{v}_d$

Symmetric matrices have orthonormal eigenvectors

So now, suppose A is symmetric:

$$\vec{v}_i^T A \vec{v}_j = \vec{v}_i^T \left(\lambda_j \vec{v}_j\right) = \lambda_j \vec{v}_i^T \vec{v}_j = \begin{cases} \lambda_j, & i = j \\ 0, & i \neq j \end{cases}$$

In other words, if a symmetric matrix has D eigenvectors with distinct eigenvalues, then its eigenvectors form an orthonormal basis:

$$\vec{v}_i^T \vec{v}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

Symmetric matrices have orthonormal eigenvectors

So now, suppose A is symmetric:

$$\vec{v}_i^T A \vec{v}_j = \vec{v}_i^T \left(\lambda_j \vec{v}_j\right) = \lambda_j \vec{v}_i^T \vec{v}_j = \begin{cases} \lambda_j, & i = j \\ 0, & i \neq j \end{cases}$$

In other words, if a symmetric matrix has D eigenvectors with distinct eigenvalues, then its eigenvectors form an orthonormal basis:

$$V^T V = I, \qquad V = \begin{bmatrix} \vec{v}_0 & \dots & \vec{v}_{D-1} \end{bmatrix}$$

Also,

$VV^T = VIV^T = VV^TVV^T = (VV^T)^2$, which is only possible if $VV^T = I$.

Eigenvectors orthogonalize a symmetric matrix:

So now, suppose A is symmetric:

$$\vec{v}_i^T A \vec{v}_j = \vec{v}_i^T \left(\lambda_j \vec{v}_j\right) = \lambda_j \vec{v}_i^T \vec{v}_j = \begin{cases} \lambda_j, & i = j \\ 0, & i \neq j \end{cases}$$

In other words, if a symmetric matrix has D eigenvectors with distinct eigenvalues, then its eigenvectors orthogonalize A:

$$V^T A V = \Lambda, \qquad \Lambda = \begin{bmatrix} \lambda_0 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & \lambda_{D-1} \end{bmatrix}$$

A symmetric matrix is the weighted sum of its eigenvectors:

So now, suppose A is symmetric:

$$\vec{v}_i^T A \vec{v}_j = \vec{v}_i^T (\lambda_j \vec{v}_j) = \lambda_j \vec{v}_i^T \vec{v}_j = \begin{cases} \lambda_j, & i = j \\ 0, & i \neq j \end{cases}$$

In other words, if a symmetric matrix has D eigenvectors with distinct eigenvalues, then it equals the weighted sum of its eigenvectors:

$$A = VV^T AVV^T = V\Lambda V^T = \sum_{d=0}^{D-1} \lambda_d \vec{v}_d \vec{v}_d^T$$

Summary: properties of symmetric matrices

If A is symmetric with D eigenvectors, and D distinct eigenvalues, then

$$A = V\Lambda V^T$$

$$\Lambda = V^T A V$$

$$VV^T = V^T V = I$$

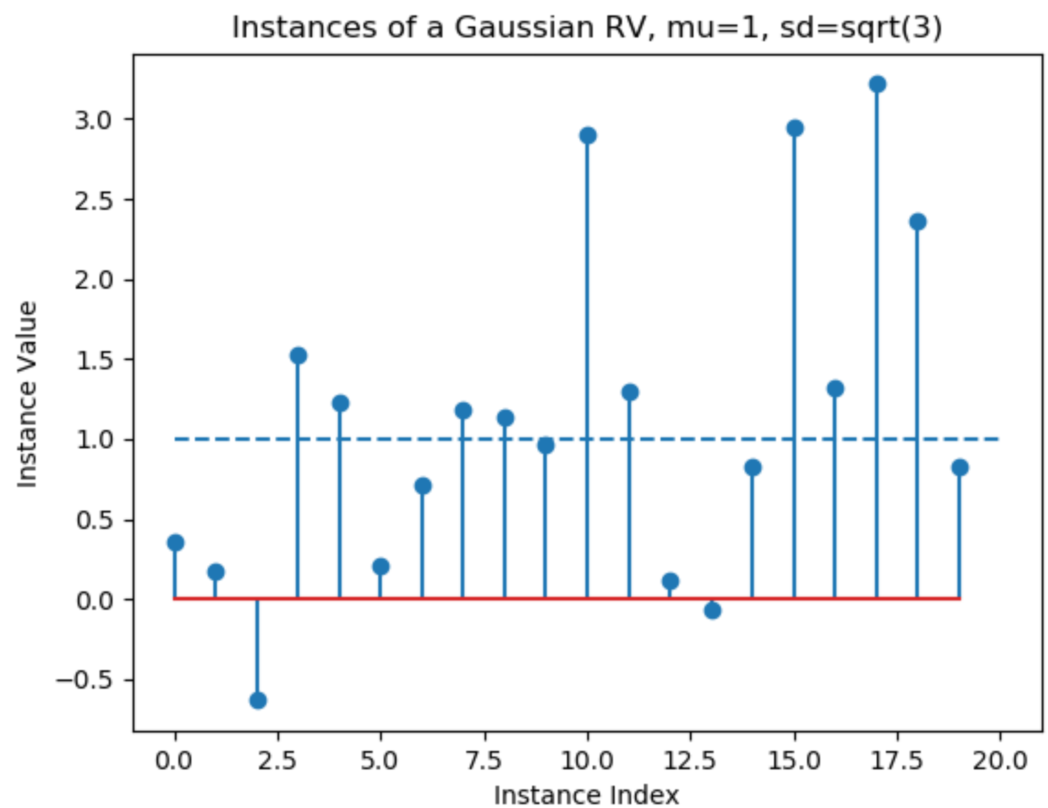
Content

- Linear transforms
- Eigenvectors
- Eigenvalues
- Symmetric matrices
- Gaussian random vectors
- Principal component axes = eigenvectors of the covariance
- Gram matrix
- Singular value decomposition

Scalar Gaussian Random Variables

$$\mu = E[x]$$

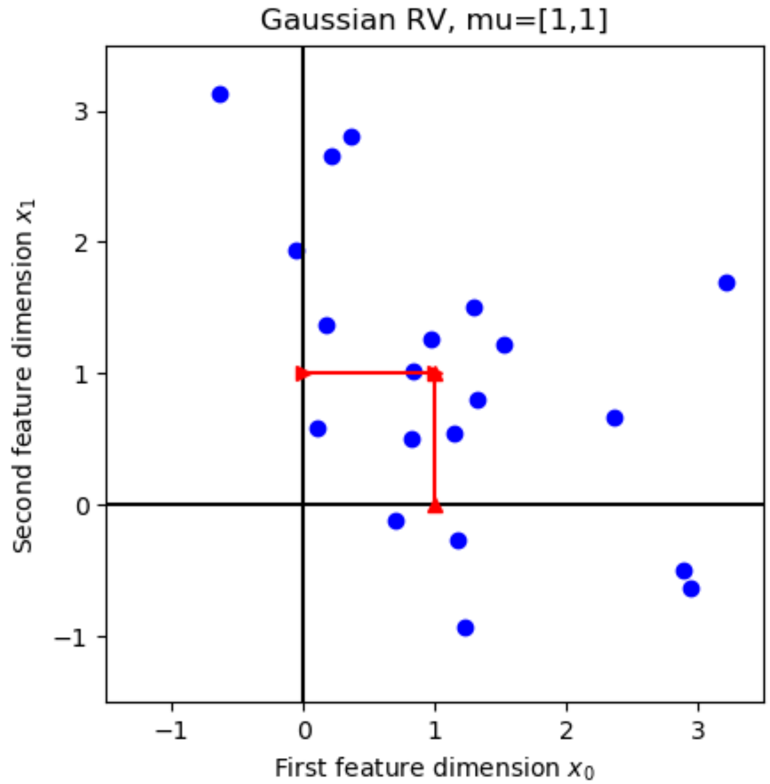
$$\sigma^2 = E[(x-\mu)^2]$$



Gaussian Random Vector

$$\vec{x} = \begin{bmatrix} x_0 \\ \vdots \\ x_{D-1} \end{bmatrix}, \vec{\mu} = \begin{bmatrix} \mu_0 \\ \vdots \\ \mu_{D-1} \end{bmatrix}$$

$$\vec{\mu} = E[\vec{x}]$$

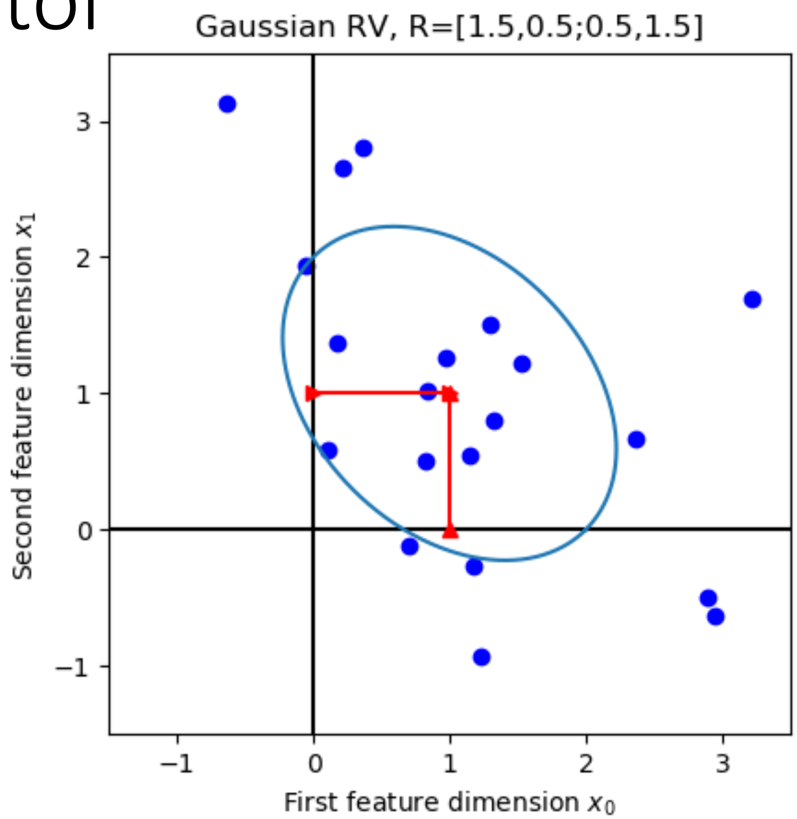


Gaussian Random Vector

$$R = \begin{bmatrix} \rho_{00} & \rho_{01} & \ddots \\ \rho_{10} & \ddots & \rho_{D-2,D-1} \\ \ddots & \rho_{D-1,D-2} & \rho_{D-1,D-1} \end{bmatrix}$$

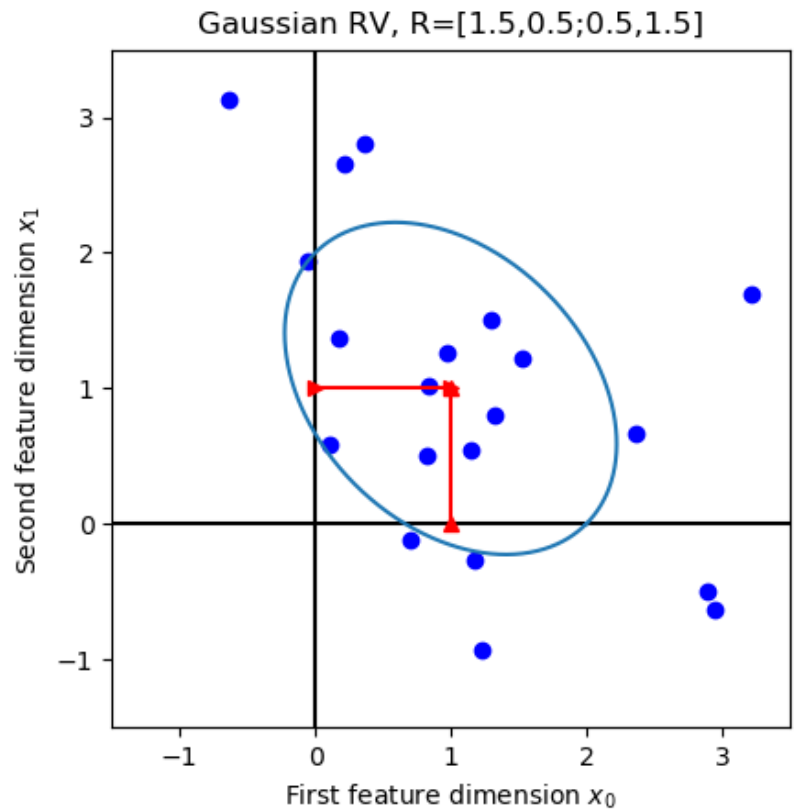
$$\rho_{ij} = E\left[(x_i - \mu_i)(x_j - \mu_j)\right]$$

$$\sigma_i^2 = \rho_{ii} = E\left[(x_i - \mu_i)^2\right]$$



Gaussian Random Vector

$$R = E[(\vec{x} - \vec{\mu})(\vec{x} - \vec{\mu})^T]$$



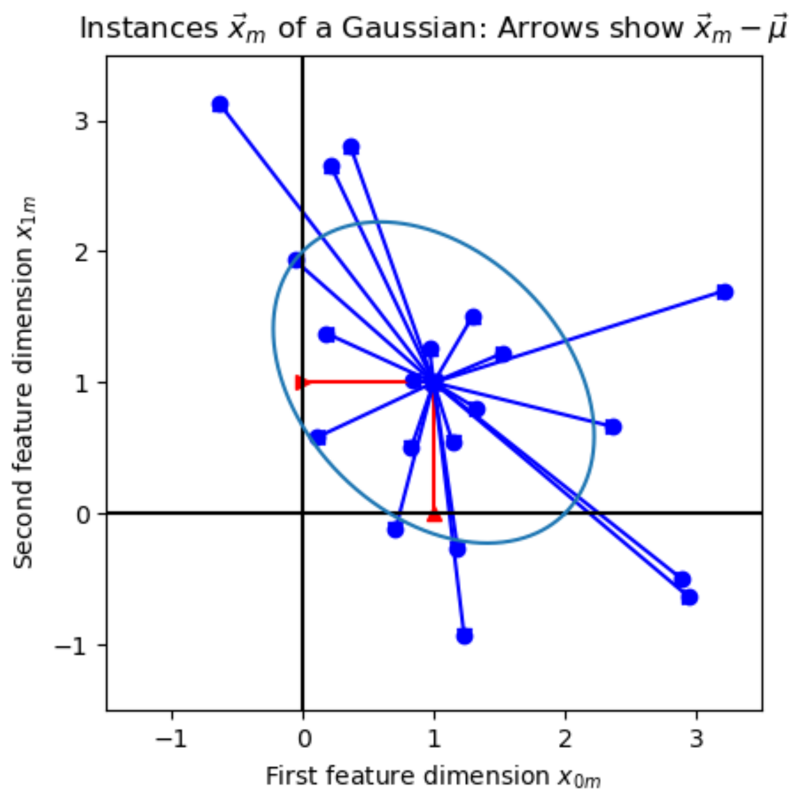
Sample Mean, Sample Covariance

In the real world, we don't know $\vec{\mu}$ and R! If we have M instances $\vec{x}_m$ of the Gaussian, we can estimate $\vec{\mu}$ and R as

$$\vec{\mu} = \frac{1}{M}\sum_{m=0}^{M-1} \vec{x}_m$$

$$R = \frac{1}{M-1}\sum_{m=0}^{M-1} (\vec{x}_m - \vec{\mu})(\vec{x}_m - \vec{\mu})^T$$

Sample mean and sample covariance are not the same as the real mean and covariance, but we'll use the same letters for them ($\vec{\mu}$ and R) unless the problem requires us to distinguish.



Content

- Linear transforms
- Eigenvectors
- Eigenvalues
- Symmetric matrices
- Gaussian random vectors
- Principal component axes = eigenvectors of the covariance
- Gram matrix
- Singular value decomposition

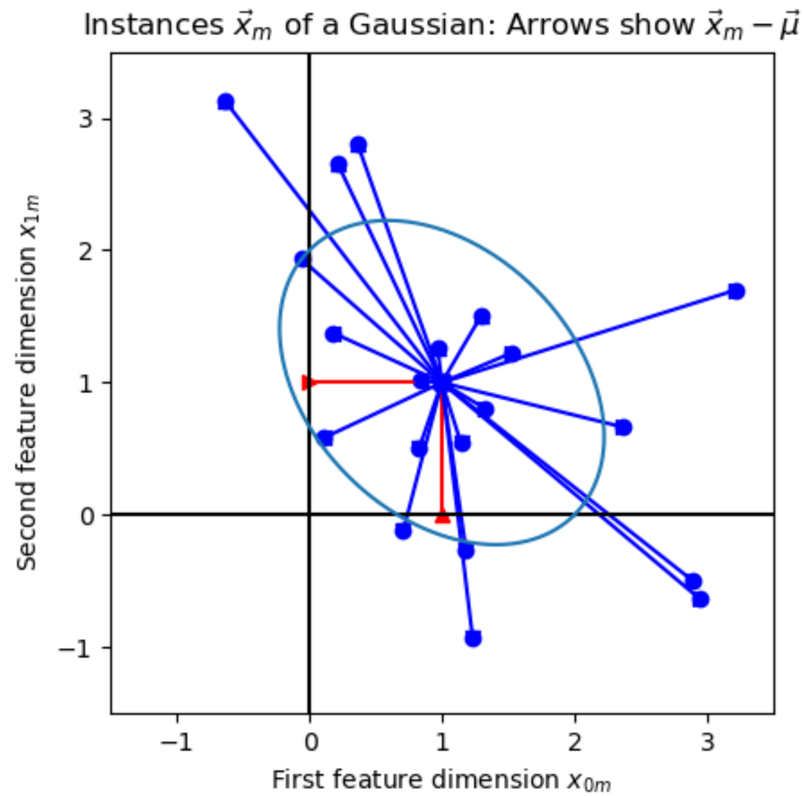
Sample Covariance

$$R = \frac{1}{M-1}\sum_{m=0}^{M-1}(\vec{x}_m - \vec{\mu})(\vec{x}_m - \vec{\mu})^T$$

$$= \frac{1}{M-1}X^T X$$

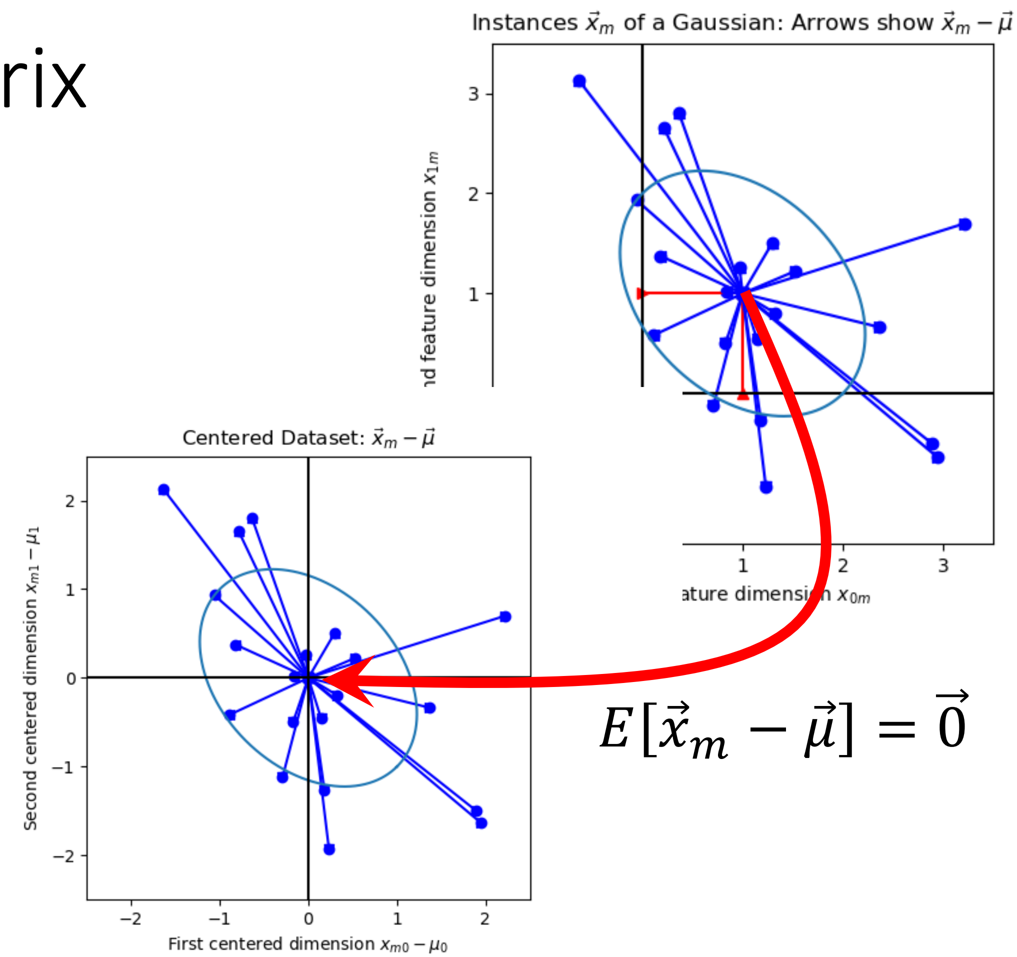
…where X is the centered data matrix,

$$X = \begin{bmatrix} (\vec{x}_0 - \vec{\mu})^T \\ \vdots \\ (\vec{x}_{M-1} - \vec{\mu})^T \end{bmatrix}$$



Centered Data Matrix

$$X = \begin{bmatrix} (\vec{x}_0 - \vec{\mu})^T \\ \vdots \\ (\vec{x}_{M-1} - \vec{\mu})^T \end{bmatrix}$$



$$E[\vec{x}_m - \vec{\mu}] = \vec{0}$$

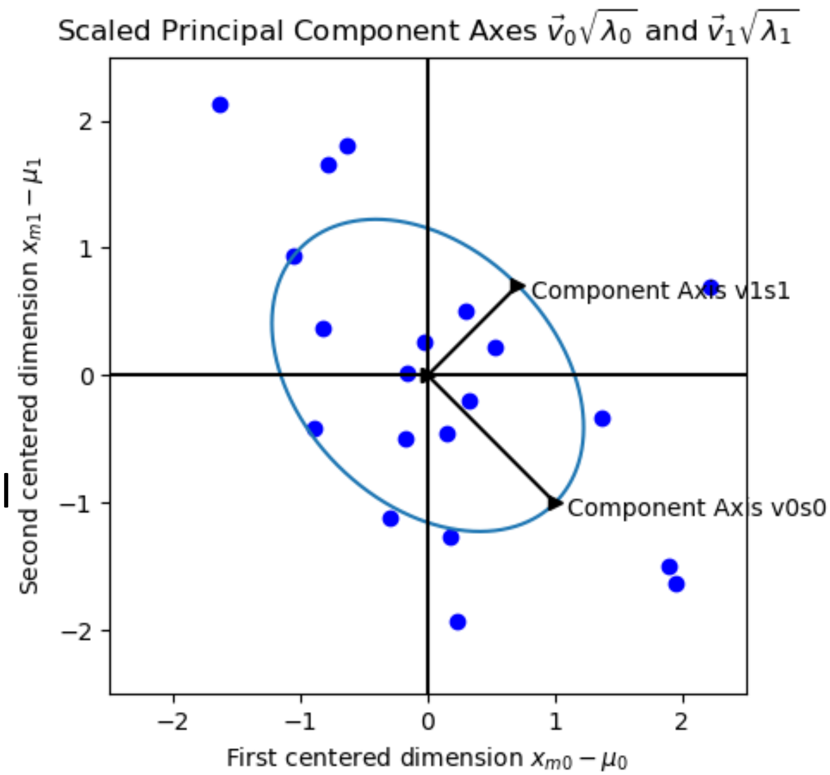
Principal Component Axes

$X^T X$ is symmetric!

Therefore,

$$X^T X = V \Lambda V^T$$

$V = [\vec{v}_0 \quad \ldots \quad \vec{v}_{D-1}]$ are called the principal component axes, or principal component directions.



Principal Components

$$X^T X = V \Lambda V^T$$

$$V^T X^T X V = V^T V \Lambda V^T V = \Lambda$$

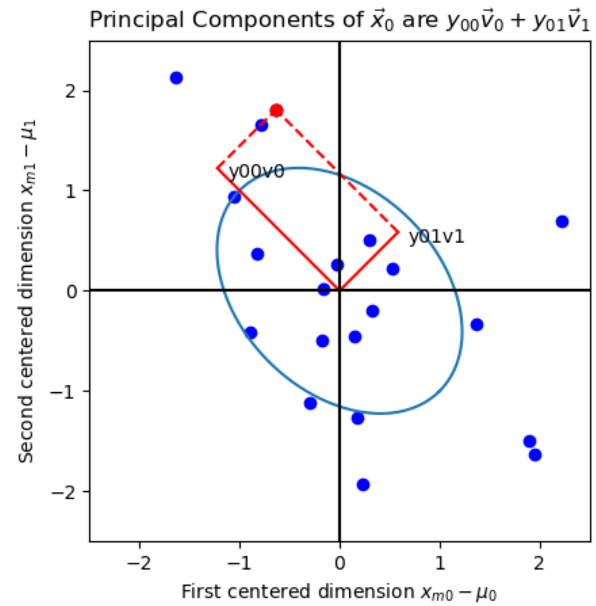
$$Y^T Y = \Lambda$$

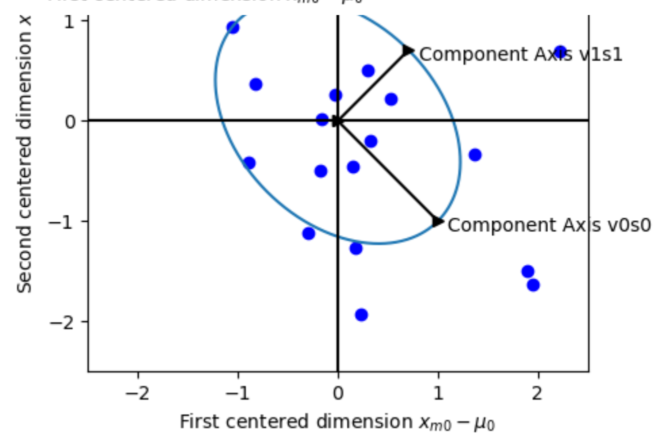
…where…

$$Y = \begin{bmatrix} \vec{y}_0^T \\ \vdots \\ \vec{y}_{D-1}^T \end{bmatrix} = \begin{bmatrix} (\vec{x}_0 - \vec{\mu})^T V \\ \vdots \\ (\vec{x}_{M-1} - \vec{\mu})^T V \end{bmatrix}$$

$\vec{y}_m$ is the vector of principal components of $\vec{x}_m$.
The principal components are:

$$y_{md} = (\vec{x}_m - \vec{\mu})^T \vec{v}_d$$

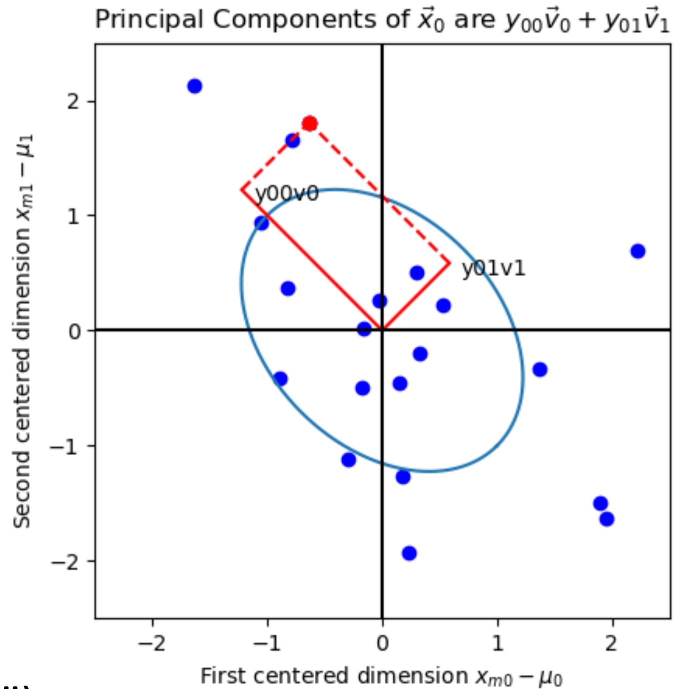




Principal Components are Uncorrelated

Basics of matrix notation. This equation: $Y^T Y = \Lambda$.
Means:

$$\sum_{m=0}^{M-1} \vec{y}_m \vec{y}_m^{\,T} = \begin{bmatrix} \lambda_0 & 0 & 0 \\ 0 & \ldots & 0 \\ 0 & 0 & \lambda_{D-1} \end{bmatrix}$$



In other words,

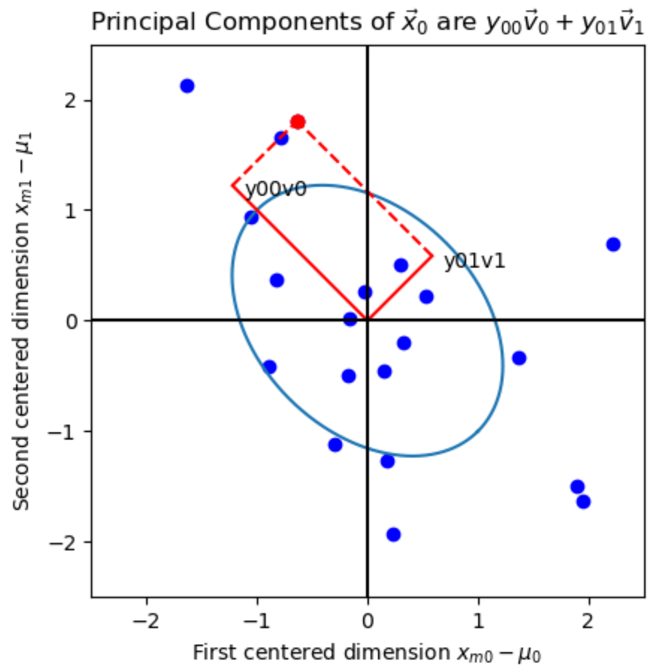
$$E\left[y_i y_j\right] \tilde{\propto} \sum_{m=0}^{M-1} y_{mi} y_{mj} = \begin{cases} \lambda_i, & i = j \\ 0, & i \neq j \end{cases}$$

(the symbol $\tilde{\propto}$ means "approximately proportional to")

Eigenvalue ∝ Variance of the Principal Component

More precisely, the sample covariance of principal components y_i and y_j is 0, unless $i = j$, in which case

$$\frac{1}{M-1}\sum_{m=0}^{M-1} y_{mi}^2 = \frac{\lambda_i}{M-1}$$



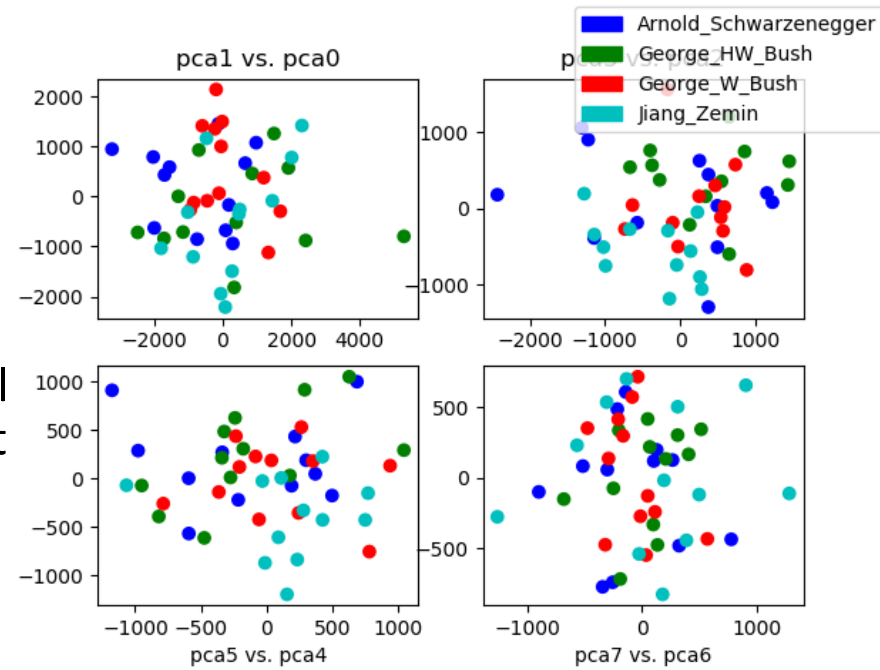
Eigenvalue = Energy of the Principal Component

The total dataset energy of y_i is

$$\sum_{m=0}^{M-1} y_{mi}^2 = \lambda_i$$

But remember that $\|\vec{v}_d\|_2 = 1$. Therefore, the total dataset energy is the same, whether you calculate it in the original image domain, or in the PCA domain:

$$\sum_{m=0}^{M-1} \sum_{d=0}^{D-1} (x_{md} - \mu_d)^2 = \sum_{m=0}^{M-1} \sum_{i=0}^{D-1} y_{mi}^2 = \sum_{i=0}^{D-1} \lambda_i$$

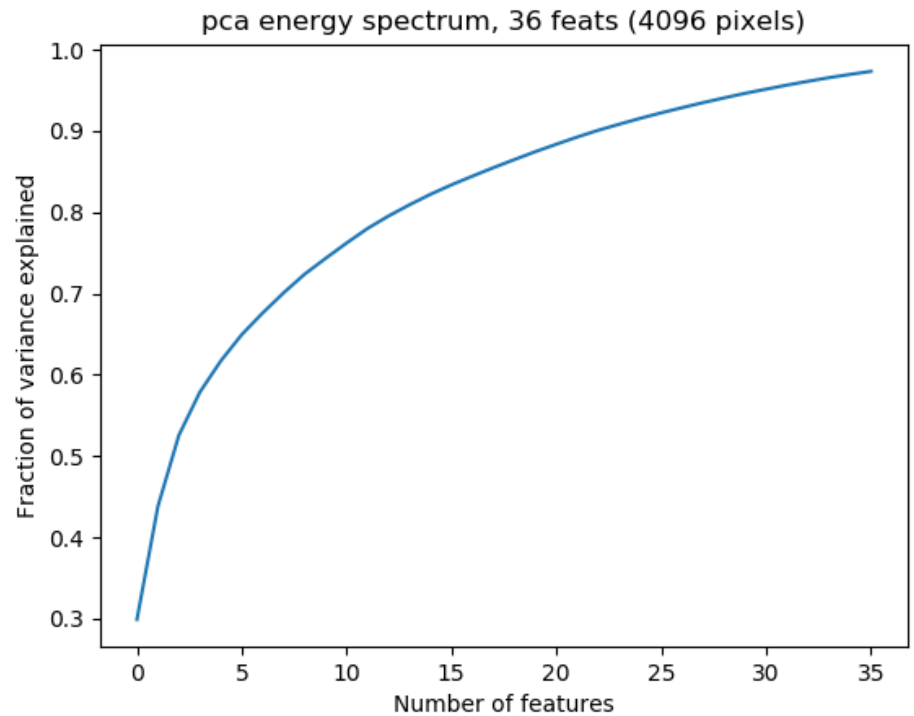


Energy Spectrum = Fraction of Energy Explained

The "energy spectrum" is energy as a function of basis vector index. There are a few ways we could define it, but one useful definition is:

$$E[k] = \frac{\sum_{m=0}^{M-1} \sum_{i=0}^{k-1} y_{mi}^2}{\sum_{m=0}^{M-1} \sum_{d=0}^{D-1} (x_{md} - \mu_d)^2}$$

$$= \frac{\sum_{i=0}^{k-1} \lambda_i}{\sum_{i=0}^{D-1} \lambda_i}$$



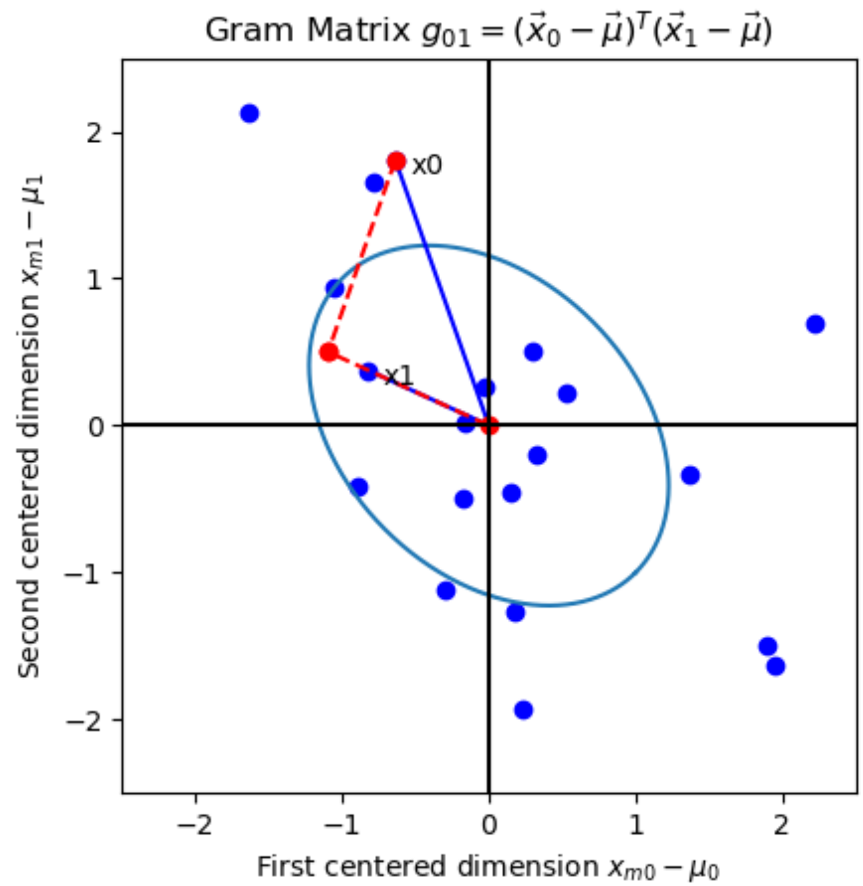
Content

- Linear transforms
- Eigenvectors
- Eigenvalues
- Symmetric matrices
- Gaussian random vectors
- Principal component axes = eigenvectors of the covariance
- Gram matrix
- Singular value decomposition

Gram Matrix

- $X^T X$ is usually called the sum-of-squares matrix. $\frac{1}{M-1} X^T X$ is the sample covariance.
- $G = XX^T$ is called the gram matrix. It's $(i,j)^{th}$ element is the dot product between the i^{th} and j^{th} data:

$$g_{ij} = (\vec{x}_i - \vec{\mu})^T \left(\vec{x}_j - \vec{\mu}\right)$$



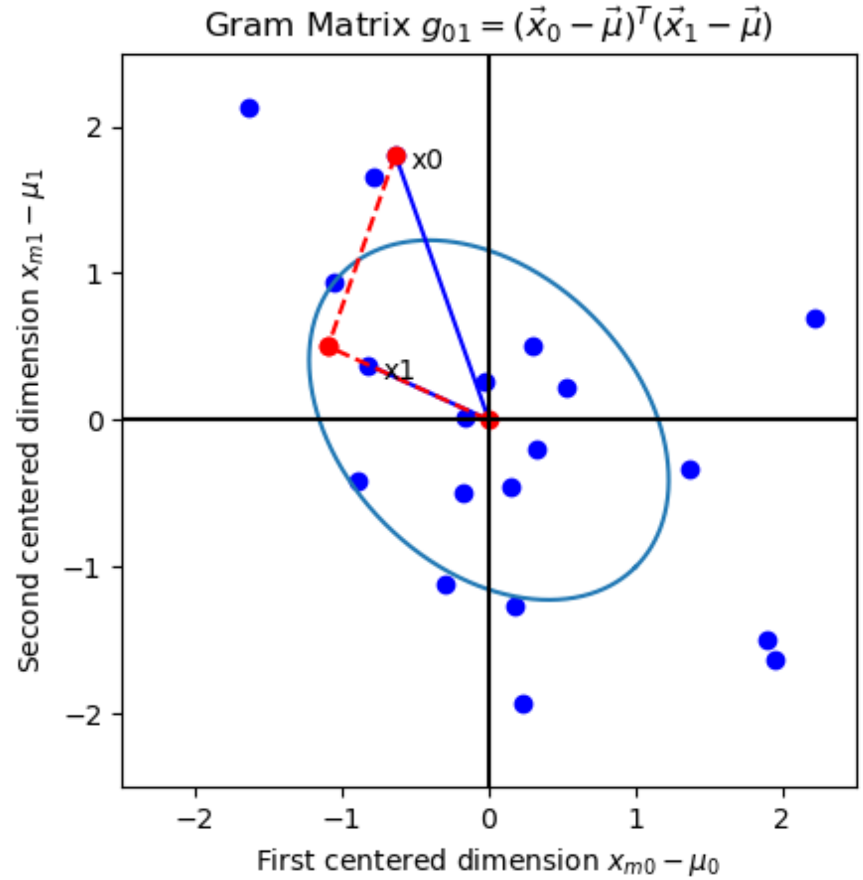
Eigenvectors of the Gram Matrix

XX^T is also symmetric! So it has orthonormal eigenvectors:

$$XX^T = U\Lambda U^T$$

$$UU^T = U^TU = I$$

X^TX and XX^T have the same eigenvalues (Λ) but different eigenvectors (V vs. U).



Why the Gram matrix is useful…

Suppose (as in mp2) that D=4096 pixels, but that M=48 images. Then, in order to perform this eigenvalue analysis:

$$X^T X = V\Lambda V^T$$

… requires factoring a 4096-order polynomial ($|X^T X - \lambda I| = 0$), then solving 4096 simultaneous linear equations in 4096 unknowns to find each eigenvector ($X^T X\vec{v}_d = \lambda_d\vec{v}_d$). Even if you use the canned function np.linalg.eig to solve it for you, it's going to take a LOT of computation. On the other hand,

$$XX^T = U\Lambda U^T$$

…is 48^{th} order. Educated experts agree: 48 < 4096.

Content

- Linear transforms
- Eigenvectors
- Eigenvalues
- Symmetric matrices
- Gaussian random vectors
- Principal component axes = eigenvectors of the covariance
- Gram matrix
- Singular value decomposition

Singular Values

- Both $X^T X$ and XX^T are positive semi-definite, meaning that their eigenvalues are non-negative, $\lambda_d \geq 0$.
- The ”Singular Values” are defined to be the square root of the eigenvalues, $s_d = \sqrt{\lambda_d}$.

$$S = \begin{bmatrix} s_0 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & s_{D-1} \end{bmatrix}, \qquad \Lambda = SS = \begin{bmatrix} s_0^2 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & s_{D-1}^2 \end{bmatrix}$$

Singular Value Decomposition

$$(M-1)R = X^T X = V\Lambda V^T = VSSV^T$$

$$G = XX^T = U\Lambda U^T = USSU^T$$

Singular Value Decomposition

$$X^T X = VSSV^T = VSISV^T = VSU^T USV^T$$

$$XX^T = USSU^T = USISU^T = USV^T VSU^T$$

Singular Value Decomposition

OK. Here is a magical fact, which is not taught in introductory linear algebra courses, and I don't know why. ANY data matrix, X, can be written as $X = USV^T$.

- $U = [\vec{u}_0 \quad \dots \quad \vec{u}_{M-1}]$ are the eigenvectors of XX^T.
- $V = [\vec{v}_0 \quad \dots \quad \vec{v}_{D-1}]$ are the eigenvectors of X^TX.
- $S = \begin{bmatrix} s_0 & 0 & 0 & 0 & 0 \\ 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & s_{\min(D,M)-1} & 0 & 0 \end{bmatrix}$ are the singular values.
    - S has some all-zero columns if M>D, or all-zero rows if M<D.

What np.linalg.svd does

First, decide whether you want to find the eigenvectors of XX^T or of X^TX: just check to see which one is larger. If you discover that 4096>48, then compute $XX^T = U\Lambda U^T$, and $S = \sqrt{\Lambda}$. Then you find V as follows:

$$X^T = VSU^T$$

$$X^TU = VS$$

$$X^TUS^{-1} = V$$

Methods that solve mp2

- Direct eigenvector analysis of $X^T X = V \Lambda V^T$ gives the right answer, but takes a very long time. When I tried this, it timed-out the autograder.
- Applying np.linalg.svd to X gives the right answer, very fast.
- This also works, and is actually just as fast as np.linalg.svd (I tested it): you can apply np.linalg.eig to the gram matrix $\mathrm{G} = XX^T$, then compute $V^T = S^{-1} U^T X$.