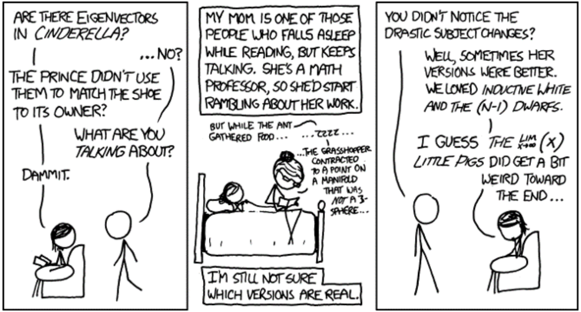
CSCI 4022 Fall 2021
Principal Component Analysis

Opening Example: Example: Find the eigenvalue/eigenvector pairs for $A = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$



Why Reduce Dimension

1. Discover hidden correlations/topics/concepts
2. Remove redundant/noisy features
3. Interpretation and visualization is easier and more intuitive in fewer dimensions
4. Easier to store, process and analyze data in fewer dimensions

Definition: $(\lambda, \boldsymbol{v})$ is an eigenpair of a matrix A if $A\boldsymbol{v} = \lambda\boldsymbol{v}$ and $v \neq 0$. So... $(A - \lambda I)\boldsymbol{v} = 0$ and since $|\boldsymbol{v}| = 1$, $|A - \lambda I| = 0$.

$\Rightarrow Av - \lambda v = 0 \Rightarrow$

(pen-and-paper) Algorithm:Write down the determinant $|A - \lambda I|$ (a polynomial) and solving for its roots.

Then, we can set up and solve the linear system $A\boldsymbol{v} = \lambda\boldsymbol{v}$ (with the restriction that $|\boldsymbol{v}| = 1$) to find the associated eigenvector.

Eigenvalues

A common step in dimension reduction is finding eigenvalues and eigenvectors of a symmetric $n \times n$ matrix. So we need an algorithm for that.

Question: What does it mean for $(\lambda, \boldsymbol{v})$ to be an eigenpair of a matrix A?

Eigenvalues

A common step in dimension reduction is finding eigenvalues and eigenvectors of a symmetric $n \times n$ matrix. So we need an algorithm for that.
Question: What does it mean for $(\lambda, \boldsymbol{v})$ to be an eigenpair of a matrix A?
Answer: It means $A\boldsymbol{v} = \lambda\boldsymbol{v}$ and $v \neq 0$

1. So... $(A - \lambda I)\boldsymbol{v} = 0$
2. Since we insist on unit eigenvectors, $|\boldsymbol{v}| = 1$
3. Result: $|A - \lambda I| = 0$ (linear algebra fact)

Eigenvalues

A common step in dimension reduction is finding eigenvalues and eigenvectors of a symmetric $n \times n$ matrix. So we need an algorithm for that.
Question: What does it mean for $(\lambda, \boldsymbol{v})$ to be an eigenpair of a matrix A?
Answer: It means $A\boldsymbol{v} = \lambda\boldsymbol{v}$ and $v \neq 0$

1. So... $(A - \lambda I)\boldsymbol{v} = 0$
2. Since we insist on unit eigenvectors, $|\boldsymbol{v}| = 1$
3. Result: $|A - \lambda I| = 0$ (linear algebra fact)

(pen-and-paper) Algorithm: So solving for eigenvalues amounts to writing down the determinant $|A - \lambda I|$ (a polynomial) and solving for its roots.

Then, we can set up and solve the linear system $A\boldsymbol{v} = \lambda\boldsymbol{v}$ (with the restriction that $|\boldsymbol{v}| = 1$) to find the associated eigenvector.

Exact Eigenvalues

1) Create $A - \lambda I = \begin{bmatrix} 3-\lambda & 2 \\ 2 & 6-\lambda \end{bmatrix}$

$I^{2\times 2}$

$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

(b) determinant $= (3-\lambda)(6-\lambda) - (2)(2)$

~3 ops

$3(3+1\ \ )$

~12 ops

Example: Find the eigenvalue/eigenvector pairs for $A = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$

set $|A - \lambda I|$

$0 = 18 - 9\lambda + \lambda^2 - 4$

$0 = \lambda^2 - 9\lambda + 14$

Exact Eigenvalues

$v = \begin{bmatrix} x \\ y \end{bmatrix}$

$\lambda = 2 \quad Av = \lambda v$

$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$

$\begin{bmatrix} 3x + 2y \\ 2x + 6y \end{bmatrix} = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$

Example: Find the eigenvalue/eigenvector pairs for $A = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$

Step 1: Characteristic Polynomial.

$$|A - \lambda I| = \left| \begin{bmatrix} 3-\lambda & 2 \\ 2 & 6-\lambda \end{bmatrix} \right| = det \begin{bmatrix} 3-\lambda & 2 \\ 2 & 6-\lambda \end{bmatrix}$$

$$= (3-\lambda)(6-\lambda) - 2 \cdot 2 \implies$$

$$0 = \lambda^2 - 9\lambda + 14 \implies \lambda = 2, 7$$

$0 = (\lambda - 2)(\lambda - 7)$

$3x + 2y = 2x$

$x = -2y$

$2x + 6y = 2y$

$x = -2y$

$x^2 + y^2 = 1$

$x = -2y \quad x^2 + y^2 = 1$

$(-2y)^2 + y^2 = 1$

$y^2 = 1/5$

$y = 1/\sqrt{5}$

$x = -2/\sqrt{5}$

(λ, v)

$\left(2, \begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}\right)$

$|v| = 1$

Exact Eigenvalues

Example: Find the eigenvalue/eigenvector pairs for $A = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$

Step 2: Eigenvectors. Using $\lambda_1 = 7$, we find v_2 via $\cancel{Av_2 = \lambda_2 v}$. Suppose $v_1 = [x, y]^T$.

$Av_1 = \lambda_1 v_1$

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 7 \begin{bmatrix} x \\ y \end{bmatrix} \implies \begin{bmatrix} 3x + 2y \\ 2x + 6y \end{bmatrix} = \begin{bmatrix} 7x \\ 7y \end{bmatrix}$$

Both rows of this system suggest that $y = 2x$, so we combine with $x^2 + y^2 = 1$ to get

$$v_1 = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

$$\left(7, \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}\right); \quad \left(2, \begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}\right)$$

Exact Eigenvalues

Example: Find the eigenvalue/eigenvector pairs for $A = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$

Final Result: The matrix A has eigenpairs of $(\lambda, \boldsymbol{v})$ of $(7, [1/\sqrt{5}, 2/\sqrt{5}]^T)$ and $(2, [2/\sqrt{5}, -1/\sqrt{5}]^T)$

Exact Eigenvalues

In practice, this is pretty awful for huge matrices, for a few reasons!

1. We often only care about *largest* eigenvalues, just like for PageRank or H/A.
2. But solving a characteristic polynomial finds *all* the eigenvalues, and it's hard to guarantee we find the largest.
3. Determinants are *very* computationally expensive

On the other hand... *we need more than just one eigenvalue*. We want **some** of the largest, but maybe not the full eigenspace. We can use *power iteration*, but **generalized power iteration** to iteratively and sequentially find eigenvalues from largest-to-smallest

Computational Eigenvalues

Idea: Find the largest eigenvalue via power iteration. Then somehow *remove* it from the matrix, so now the *second highest* has become the highest. Repeat power iteration!

Recall: Power Iteration

Computational Eigenvalues

Idea: Find the largest eigenvalue via power iteration. Then somehow *remove* it from the matrix, so now the *second highest* has become the highest. Repeat power iteration!

Recall: Power Iteration

Let A be a symmetric $n \times n$ matrix, whose eigenstuff we want to find. Start with some $x_0 \neq 0$, then

$x^{new} = Ax^{old}$; then $x^{new} = x^{new} / |x^{new}|$

1. Iterate until convergence: $||x_k - x_{k+1}||_F < tol$
2. Then our final x is the *principal* eigenvector of A, x_1, and we can solve for the associated eigenvalue λ_1 via:

   $x^T x = \sum x_i^2 = 1$

$$\boxed{Ax = \lambda x} \implies x^T Ax = x^T \lambda x \implies \lambda = x^T Ax$$

Note: the matrix/vector norm used there is the Frobenius norm:

$$||A_F|| = \sqrt{\sum_{i,j} A_{i,j}^2}$$

Computational Eigenvalues

$\lambda = 7$ $\quad x = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$

$A = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$

$x x^T = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 1/5 & 2/5 \\ 2/5 & 4/5 \end{bmatrix}$

How do we remove an eigenvalue to set up finding the second-biggest?

Process: Set

$$A_2 = A - \lambda_1 x_1 x_1^T$$

$A_2 = \begin{bmatrix} 3 - 7/5 & 2 - 7 \cdot 2/5 \\ 2 - 7 \cdot 2/5 & 6 - 7 \cdot 4/5 \end{bmatrix}$

Computational Eigenvalues

How do we remove an eigenvalue to set up finding the second-biggest?
Process: Set

$$A_2 = A - \lambda_1 \underbrace{x_1 x_1^T}_{\text{matrix with (i,j) component} x_i x_j}$$

...and then do power method!

1. Iterate exactly the same way that led you to x_1 to find x_2, the eigenvector associated with the second-largest eigenvalue.
2. Solve for the second eigenvalue the same way too, as: $\lambda_2 = x_2^T A x_2$
3. and continue until you found them all (or however many you wanted), updating the matrix as

$$A_{k+1} = A_k - \lambda_k x_k x_k^T$$

Computational Eigenvalues

A evals $(7,2) \Rightarrow A_2$ evals $(2, 0)$

How do we remove an eigenvalue to set up finding the second-biggest?

Process: Set

$$A_2 = A - \lambda_1 \underbrace{x_1 x_1^T}_{\text{matrix with (i,j) component} x_i x_j}$$

"remove the e.val λ_1 from A"

...and then do power method!

1. Iterate exactly the same way that led you to x_1 to find x_2, the eigenvector associated with the second-largest eigenvalue.
2. Solve for the second eigenvalue the same way too, as: $\lambda_2 = x_2^T A x_2$
3. and continue until you found them all (or however many you wanted), updating the matrix as

$$A_{k+1} = A_k - \lambda_k x_k x_k^T$$

Theory for Power Iteration

Can we convince ourselves that $A_2 = A - \lambda_1 x_1 x_1^T$ "removes" the eigenvector x_1 from A? What does that even mean?

Proof: Suppose (λ, x) is an eigenpair of matrix A, and $x \neq x_1$ (the principal eigenvector).

Claim: (λ, x) is an eigenpair of A_2.

s'pose $Ax_2 = \lambda_2 x_2$

Proof: $A_2 x = (A - \lambda_1 x_1 x_1^T) x = Ax - \lambda_1 x_1 x_1^T x = \lambda x - \lambda_1 (0) = \lambda x$ ✓

know

Claim: x_1 is also an eigenvector of A_2, but its corresponding eigenvalue is 0. $Ax_1 = \lambda_1 x_1$

Proof: $A_2 x_1 = (A - \lambda_1 x_1 x_1^T) x_1 = Ax_1 - \lambda_1 x_1 x_1^T x_1 = \lambda_1 x_1 - \lambda_1 x_1 (1) = 0 = 0 x_1$ ✓

$A_2 x_1 \overset{?}{=} 0 x_1$ $(A - \lambda_1 x_1 x_1^T) \cdot x_1$

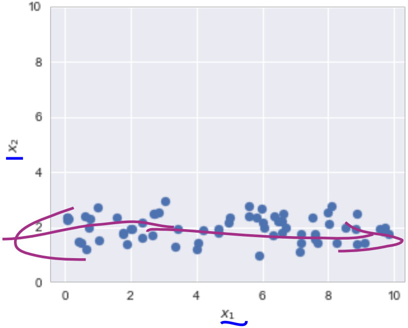
Principal Component Analysis

So what can we do with an eigendecomposition of a matrix? *Just about anything!*

Big idea (PCA): Take a data set of tuples (generally in high-dimensional space) and find the directions (or axes) along which the tuples line up best.

Example 1: x_1 is the most "important" direction to describe a data point here.

also known as the vectorized direction $[1, 0]$



Principal Component Analysis

So what can we do with an eigendecomposition of a matrix? *Just about anything!*

Big idea (PCA): Take a data set of tuples (generally in high-dimensional space) and find the directions (or axes) along which the tuples line up best.

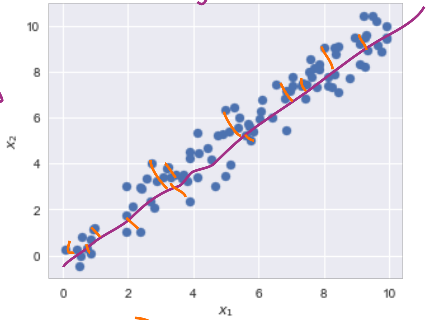
y=x?

Example 1: x_1 is the most "important" direction to describe a data point here.

[1]

also known as the vectorized direction $[1, 0]$

Example 2: What about here?



error?

Principal Component Analysis

So what can we do with an eigendecomposition of a matrix? *Just about anything!*

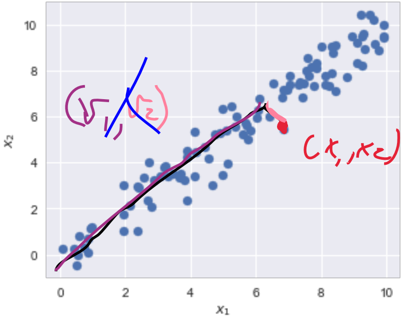
Big idea (PCA): Take a data set of tuples (generally in high-dimensional space) and find the directions (or axes) along which the tuples line up best.

Example 1: x_1 is the most "important" direction to describe a data point here.

also known as the vectorized direction $[1, 0]$

Example 2: What about here?

Now the direction of $[1, 1]$ is most "important"



The goal: Figure out how we compute which direction(s) are the most important.

These are the *principal components*

Opening

PCA overview

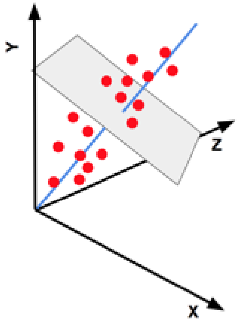
$\underset{n \times D}{M} \underset{D \times n}{M^T} = n \times n$ $\qquad$ $\underset{D \times n}{M^T} \underset{n \times D}{M} = D \times D$

Game plan:

1. Line up the data point tuples as rows of a matrix M ($n \times D$, where $n = \#$ data and $D = \#$ dimensions) $(D < n$, normally$)$
2. Find the eigenpairs of MM^T or M^TM
   (What are the sizes?)
3. Principal eigenvector, v_1, is the axis along which the points are spread out the most (highest variance)
4. The second eigenvector, v_2 (associated with the second-largest eigenvalue), is the axis along which the variance of the points' distances from the first axis (v_1) is greatest
   Note that this is always an orthogonal direction, since they're eigenvectors!
5. . . . and so on. . .

PCA is Data Mining

1. High-dimensional data can be replaced by its projection onto only the most important axes (the principal components)
2. Those correspond to the largest eigenvalues
3. So the original data can be summarized by / approximated by data with fewer dimensions – only keep what matters!



PCA; Example

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

$MM^T \to (4\times 2)\times(2\times 4) = 4\times 4$

$M^TM \to 2\times 2$

We'll walk through a stylized and simplistic example to illustrate how PCA is done.

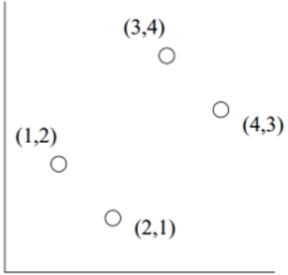
Example: Consider the set of data points (1,2), (2,1), (3,4), (4,3).

Step 1:

1. Construct M, and M^TM (or MM^T, whichever is smaller)

Solution:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 30 & 28 \\ 28 & 30 \end{bmatrix}$$



Theorem: For any real matrix A, A^TA is symmetric.

PCA; Example

We'll walk through a stylized and simplistic example to illustrate how PCA is done.

Example: Consider the set of data points (1,2), (2,1), (3,4), (4,3).

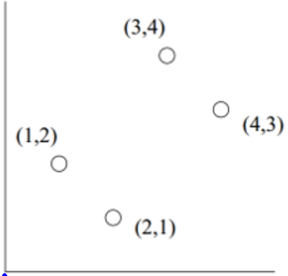
Step 1:

1. Construct M, and $M^T M$ (or $M M^T$, whichever is smaller)

Solution:

$$M = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix}; \quad M^T M = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 30 & 28 \\ 28 & 30 \end{bmatrix}$$

coincidence



Theorem: For any real matrix A, $A^T A$ is symmetric.

always equal

PCA; Example

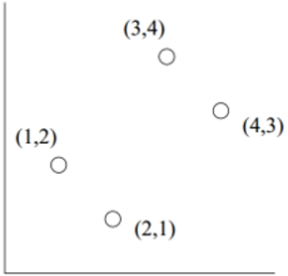
Example: Consider the set of data points (1,2), (2,1), (3,4), (4,3).

We have that $M^T M = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 30 & 28 \\ 28 & 30 \end{bmatrix}$

Step 2:

2. Find the eigenpairs for $A^T A$.

Solution:



PCA; Example

Example: Consider the set of data points (1,2), (2,1), (3,4), (4,3).

We have that $M^TM = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 30 & 28 \\ 28 & 30 \end{bmatrix}$

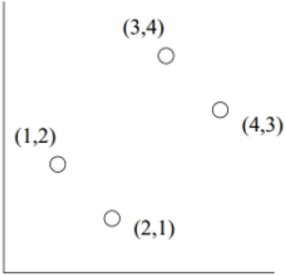
Step 2:

2. Find the eigenpairs for A^TA.

Solution:

$$|M^TM - \lambda I| = | \begin{bmatrix} 30 & 28 \\ 28 & 30 \end{bmatrix} | = (30 - \lambda)^2 - 28^2$$

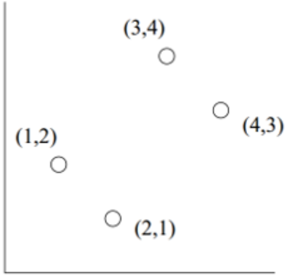
which has eigenvalues of $\lambda_1 = 58$, $\lambda_2 = 2$.



PCA; Example

Example: Consider the set of data points (1,2), (2,1), (3,4), (4,3).

We have that $\lambda_1 = 58$, $\lambda_2 = 2$
Step 2b:
2b. Find the eigenvectors for $A^T A$.
Solution:

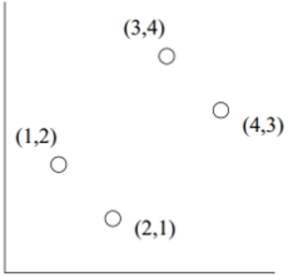


PCA; Example

Example: Consider the set of data points (1,2), (2,1), (3,4), (4,3).

We have that $\lambda_1 = 58$, $\lambda_2 = 2$
Step 2b:
2b. Find the eigenvectors for $A^T A$.
Solution: For $\lambda_1 = 58$:

$$M^T M v_1 = \begin{bmatrix} 30 & 28 \\ 28 & 30 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \overset{\heartsuit}{=} 58 \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$



PCA; Example

Example: Consider the set of data points (1,2), (2,1), (3,4), (4,3).

We have that $\lambda_1 = 58$, $\lambda_2 = 2$

Step 2b:

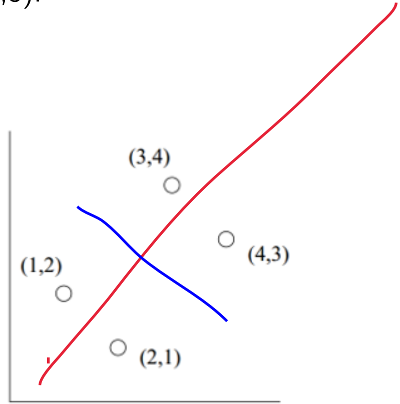
2b. Find the eigenvectors for $A^T A$.

Solution: For $\lambda_1 = 58$:

$$M^T M v_1 = \begin{bmatrix} 30 & 28 \\ 28 & 30 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \overset{\heartsuit}{=} 58 \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

so $30v_1 + 28v_2 = 58v_1$, which has solution of $v_1 = v_2$ and normalizes to $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$

Similarly, $v_2 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$



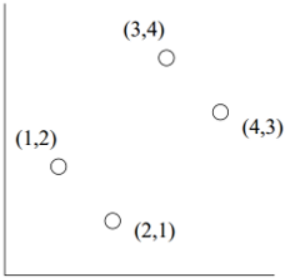
The Rotation

So we have the eigenpairs of $M^T M$.

Step 3:

3. Step 3. Construct E, the matrix whose *columns* are made up of the eigenvectors.

Solution:



The Rotation

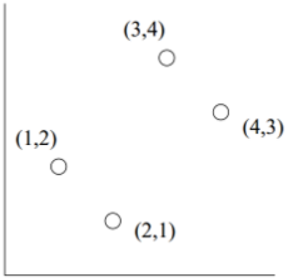
So we have the eigenpairs of $M^T M$.

Step 3:

3. Step 3. Construct E, the matrix whose *columns* are made up of the eigenvectors.

Solution:

$$E = \begin{bmatrix} \uparrow & \uparrow \\ v_1 & v_2 \\ \downarrow & \downarrow \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$



The Rotation

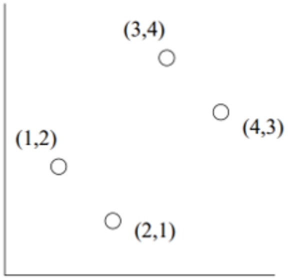
So we have the eigenpairs of $M^T M$.

Step 3:

3. Step 3. Construct E, the matrix whose *columns* are made up of the eigenvectors.

Solution:

$$E = \begin{bmatrix} \uparrow & \uparrow \\ v_1 & v_2 \\ \downarrow & \downarrow \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$



Definition: Any matrix - such as E - whose columns are made up of **orthonormal vectors** represents a **rotation** or **reflection** in a Euclidean space.

Example: $T = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is the 2-D matrix that rotates clockwise by θ

Rotations

How does that work?

Example: Example: Consider the row vector $x = [0, 1]$ in $\mathbb{R}^2$. What is the matrix T that will rotate it clockwise by $90°$? Perform this transformation xT to find the resulting rotated vector.

Solution:

Rotations

How does that work?

Example: Example: Consider the row vector $x = [0, 1]$ in $\mathbb{R}^2$. What is the matrix T that will rotate it clockwise by $90°$? Perform this transformation xT to find the resulting rotated vector.

Solution:

$$T = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \cos\pi/2 & -\sin\pi/2 \\ \sin\pi/2 & \cos\pi/2 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

so $xT = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix}$ And [1, 0] is indeed what we expect if we were to rotate $[0, 1]$ clockwise by $90°$. Hooray!

Example, cont'd: Our rotation matrix is $E = \begin{bmatrix} \uparrow & \uparrow \\ v_1 & v_2 \\ \downarrow & \downarrow \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

Rotations

How does that work?

Example: Example: Consider the row vector $x = [0, 1]$ in $\mathbb{R}^2$. What is the matrix T that will rotate it clockwise by $90°$? Perform this transformation xT to find the resulting rotated vector.

Solution:

$$T = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \cos\pi/2 & -\sin\pi/2 \\ \sin\pi/2 & \cos\pi/2 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

so $xT = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix}$ And [1, 0] is indeed what we expect if we were to rotate $[0, 1]$ clockwise by $90°$. Hooray!

Example, cont'd: Our rotation matrix is $E = \begin{bmatrix} \uparrow & \uparrow \\ v_1 & v_2 \\ \downarrow & \downarrow \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$, so expect it will rotate our original data points by $45°$ (since $\cos^{-1}(1/\sqrt{2}) = \pi/4$.)

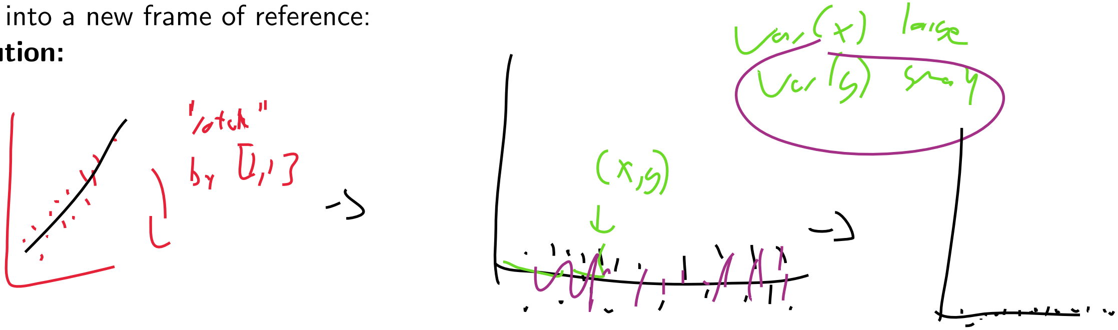
Rotations and PCA

So we have the *rotation* represented by $E = \begin{bmatrix} \uparrow & \uparrow \\ v_1 & v_2 \\ \downarrow & \downarrow \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

Step 4:

4. Rotate the original data points M by E. This puts them into a new frame of reference:

Solution:



Rotations and PCA

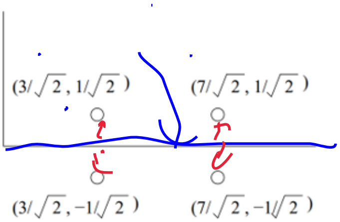
So we have the *rotation* represented by $E = \begin{bmatrix} \uparrow & \uparrow \\ v_1 & v_2 \\ \downarrow & \downarrow \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

Step 4:

4. Rotate the original data points M by E. This puts them into a new frame of reference:

Solution:

$$ME = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 3/\sqrt{2} & 1/\sqrt{2} \\ 3/\sqrt{2} & -1/\sqrt{2} \\ 7/\sqrt{2} & 1/\sqrt{2} \\ 7/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$



Rotations and PCA

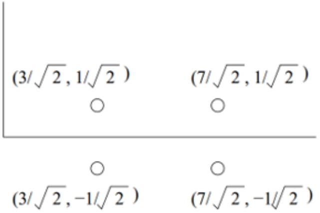
So we have the *rotation* represented by $E = \begin{bmatrix} \uparrow & \uparrow \\ v_1 & v_2 \\ \downarrow & \downarrow \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

Step 4:

4. Rotate the original data points M by E. This puts them into a new frame of reference:

Solution:

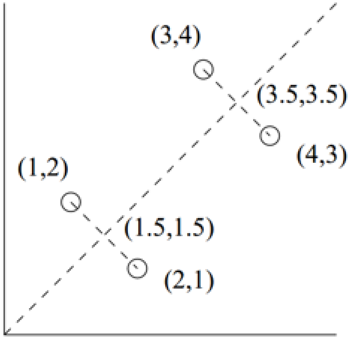
$$ME = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 3/\sqrt{2} & 1/\sqrt{2} \\ 3/\sqrt{2} & -1/\sqrt{2} \\ 7/\sqrt{2} & 1/\sqrt{2} \\ 7/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$



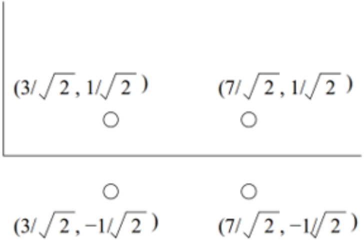
Upshot: The rows of ME give the coordinates in the eigenvector basis (along the principal components) of the corresponding rows (data points) in M.

Rotations and PCA

$$ME = \begin{bmatrix} 3/\sqrt{2} & 1/\sqrt{2} \\ 3/\sqrt{2} & -1/\sqrt{2} \\ 7/\sqrt{2} & 1/\sqrt{2} \\ 7/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$ *rotates* M into a new frame of reference.



Rotates to



Sanity Check: Does the geometry work out?

Rotations and PCA

$$ME = \begin{bmatrix} 3/\sqrt{2} & 1/\sqrt{2} \\ 3/\sqrt{2} & -1/\sqrt{2} \\ 7/\sqrt{2} & 1/\sqrt{2} \\ 7/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

The rows of ME give the coordinates in the **eigenvector basis** (along the principal components) of the corresponding rows (data points) in M.

In other words, if we took $3/\sqrt{2}x_1 + 1/\sqrt{2}x_2$, we'd have a new location for datum $\#1$.

Interpretation: The components in ME give the distance to each data point along the eigenvectors.

...and reduce

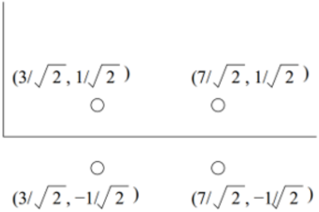
So we have the *projection* represented by ME... but let's not use all of it!

Step 5:

5. To get a reduced dimension representation of our data in M, can take only the eigenvectors associated with the k largest eigenvalues of M^TM (or MM^T).

Put them in a *narrower* matrix, E_k.

Then ME_k is a reduced-dimension representation of M, in the directions of the



Back to the running **example**: There are only $D = 2$ dimensions, so the only reducing we can do is to $d = 1$. So take the principal eigenvector and reduce:

$$ME_1 = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 3/\sqrt{2} \\ 3/\sqrt{2} \\ 7/\sqrt{2} \\ 7/\sqrt{2} \end{bmatrix}$$

... is an *approximation* of M with less columns! How close is it?

Why bother?

There are two algorithms we've seen that PCA plays well with:

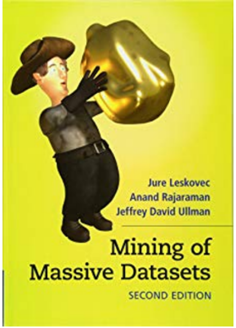
1. Dimension reduction that encourages orthonormal axes for the principal components helps with multiple linear regression (and logistic variants).
   TL;DR: if feature/columns are similar/dependent, then MLR inference fails due to $\hat{\beta} \propto (X^T X)^{-1}$, which is singular if columns of X are identical to one another.
2. Clustering is more numerically stable with lower dimension... it also visualizes better! See also: multidimensional scaling.
   TL;DR: in too many dimensions, distances tend to all look similar.

One thing to note: our "approximation" of M was a truncated *rotation* $ME \to ME_1$. It might make sense to rotate that approximation *back* into the original units of M, since ME_1 has been rotated into the frame of references of that original largest component. That way, we can actually directly compare M to it's lower-dimensional representation! (We do this next time!)

Acknowledgments

Next time: *singular* matrix decompositions

Some material is adapted/adopted from Mining of Massive Data Sets, by Jure Leskovec, Anand Rajaraman, Jeff Ullman (Stanford University) `http://www.mmds.org`



Special thanks to Tony Wong for sharing his original adaptation and adoption of slide material.