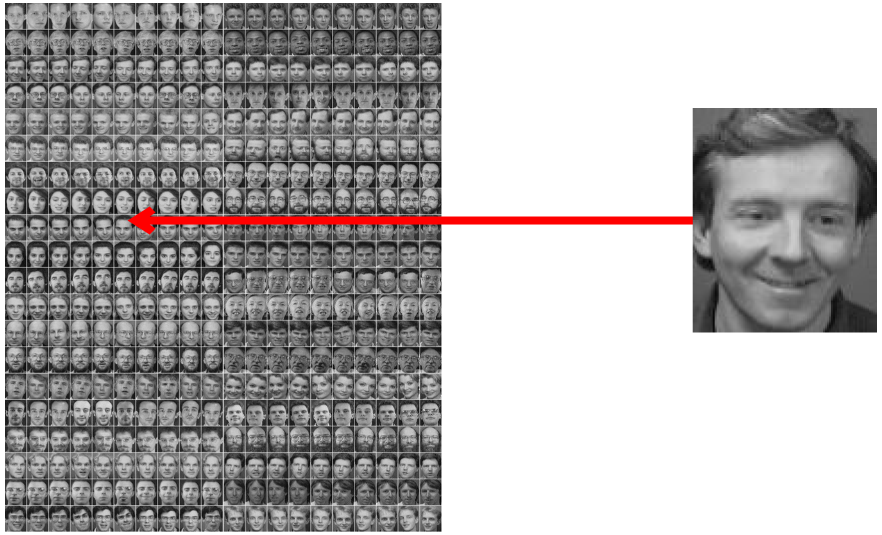
Foundational Math Data Science
Sharat Chandran

(Many slides obtained from colleagues, and gratefully acknowledged)

Face Recognition

- Given a database of face images of people, and a new *cropped* test image answer: “Who is this person?
  - Subtle point: what if the cropped test image is not that of a person?



Face Recognition System

(1) Collect database of face images of people. Record one or multiple images per person (called **"gallery image(s)"** of the person)

(2) Normalize the images: crop face from the image background

(3) Extract relevant features from the normalized face image

(4) Map the features extracted from the image with corresponding person, and store in a database

TRAINING (Offline)

Face Recognition System

(1) Alignment: Normalize the **probe image**

(3) Find the gallery image whose features most closely match (**nearest neighbor search**) those of the probe image.

(2) Extract features (as was done before for the gallery images)

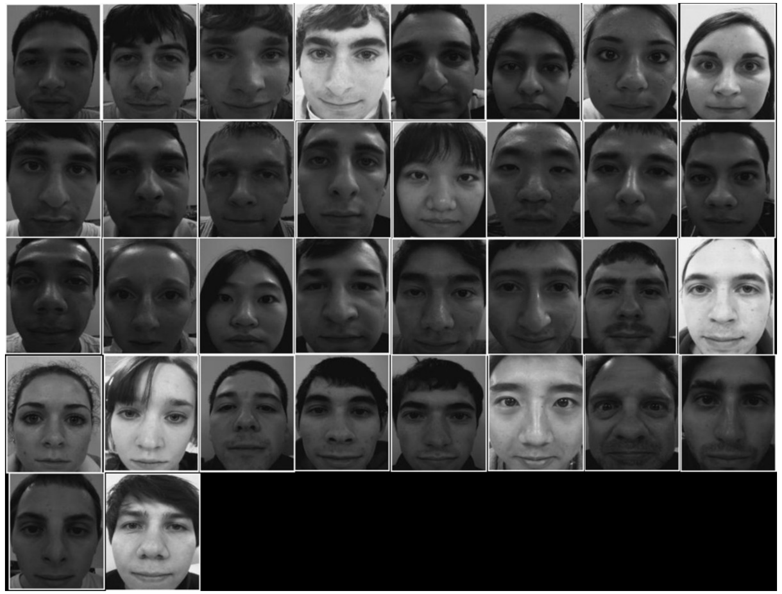
*For now, we assume that the person whose identity was to be determined, has images recorded in the gallery database

TESTING

Face Recognition System

(1) Collect database of face images of people. Record one or multiple images per person (called **"gallery image(s)"** of the person)

(2) Normalize the images: crop face from the image background

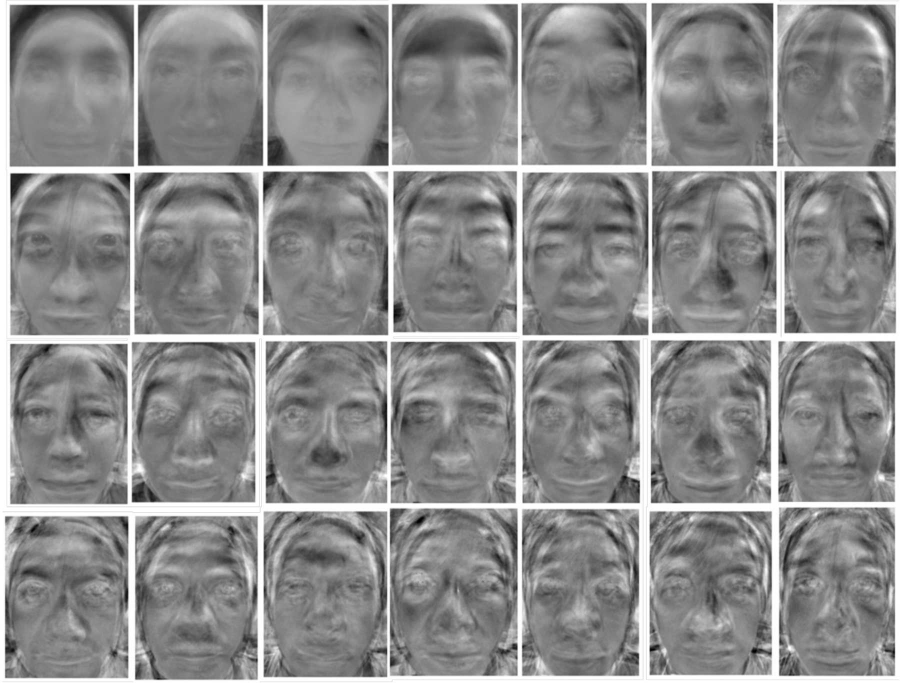


34 faces

In this example we see cropped faces.
We are going to reshape each of these into vectors each of whose dimension is height x width

TRAINING

Face Recognition System



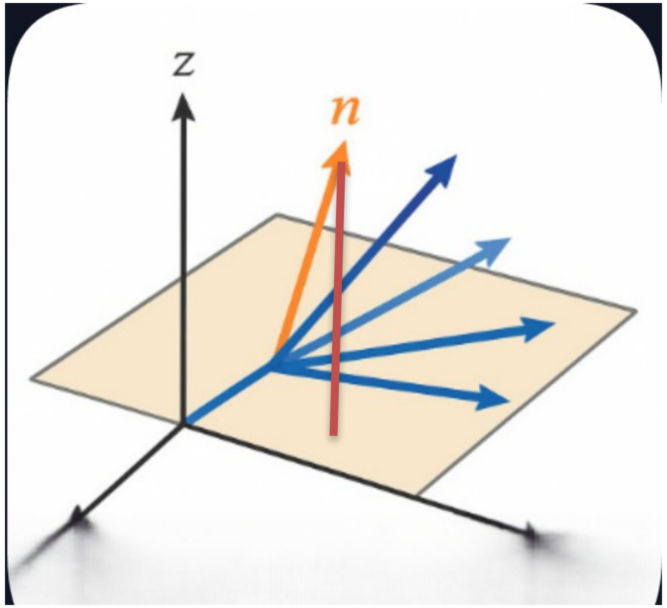
(4) Map the features extracted from the image with corresponding person, and store in a database

"Relevant" features are 28 vectors (regardless of the number of persons in the database)
Each of the 34 persons is identified by a set of 28 scalars

(3) Extract relevant features from the normalized face image

TRAINING

Face Recognition System



Project *n* onto the space formed by the 28 vectors to get 28 numbers for this probe image

How to project?

(2) Extract features for query image (as was done before for the gallery images)

TESTING

Digression

Matrix Multiplication

- Concept #1: Matrix multiplication is row times a column
  - Traditional definition of matrix multiplication
  - A bunch of dot products
  - Also called the inner product

Matrix Multiplication

- Concept #1: MM is row times column
- Concept #2: Matrix multiplication is column times row
  - The multiplication is repeated sums (of column times rows)
  - outer product

$$\mathbf{A} = [a_{ij}]_{m\times n}\mathrm{B}{=}[\mathrm{b}_{ij}]_{n\times p}\mathrm{C}{=}\mathrm{AB}{=}[\mathrm{c}_{ij}]_{m\times p}.$$

$$c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$$

$$\mathbf{C} = \begin{pmatrix} \sum_{k=1}^{n} a_{1k}b_{k1} & \sum_{k=1}^{n} a_{1k}b_{k2} & \cdots & \sum_{k=1}^{n} a_{1k}b_{kp} \\ \sum_{k=1}^{n} a_{2k}b_{k1} & \sum_{k=1}^{n} a_{2k}b_{k2} & \cdots & \sum_{k=1}^{n} a_{2k}b_{kp} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^{n} a_{mk}b_{k1} & \sum_{k=1}^{n} a_{mk}b_{k2} & \cdots & \sum_{k=1}^{n} a_{mk}b_{kp} \end{pmatrix}$$

$$= \sum_{k=1}^{n} \begin{pmatrix} a_{1k}b_{k1} & a_{1k}b_{k2} & \dots & a_{1k}b_{kp} \\ a_{2k}b_{k1} & a_{2k}b_{k2} & \dots & a_{2k}b_{kp} \\ \vdots & \vdots & \ddots & \vdots \\ a_{mk}b_{k1} & a_{mk}b_{k2} & \dots & a_{mk}b_{kp} \end{pmatrix}$$

$$= \sum_{k=1}^{n} \begin{pmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{pmatrix} \begin{pmatrix} b_{k1} & b_{k2} & \dots & b_{kp} \end{pmatrix}$$

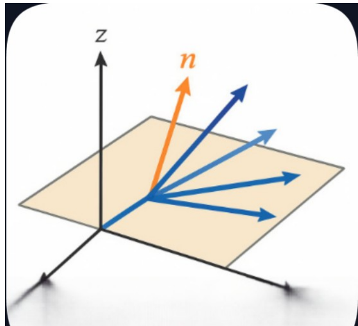
Matrix Multiplication

- Concept #1: Row times column (inner product)
- Concept #2: Repeated sums of column times row (outer product)
- Concept #3: AB = A [**b1 | b2**] = [**Ab1 | Ab2**]

End Digression

Projection: Best x when $Ax \neq b$

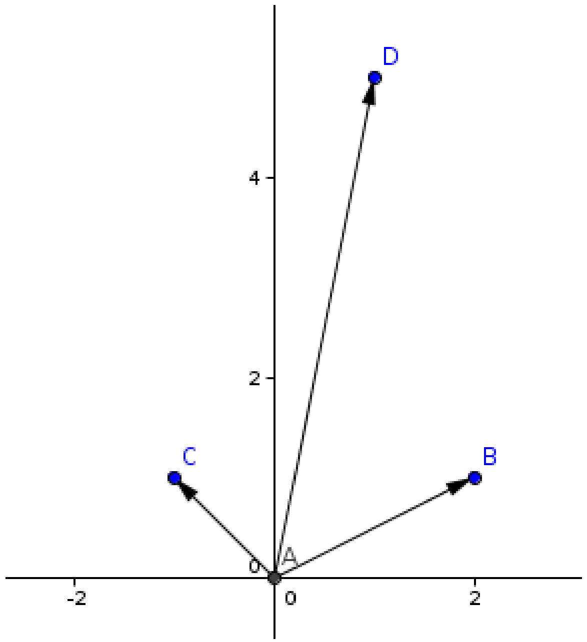
- With n linearly independent vectors (a_1, a_2, ... a_n) in R^m, find the combination $p = w_1 a_1 + ... + w_n a_n$ closest to a given vector b
- Solution: The error vector $b - Aw$ is perpendicular to the subspace, and thus to each of the n vectors a_i
  - Write $a_i^T (b - Aw) = 0$ implies $A^T A w = A^T b$
  - If the matrix $A^T A$ is invertible, then we can find w and thus p
  - How to tell if $A^T A$ is invertible?

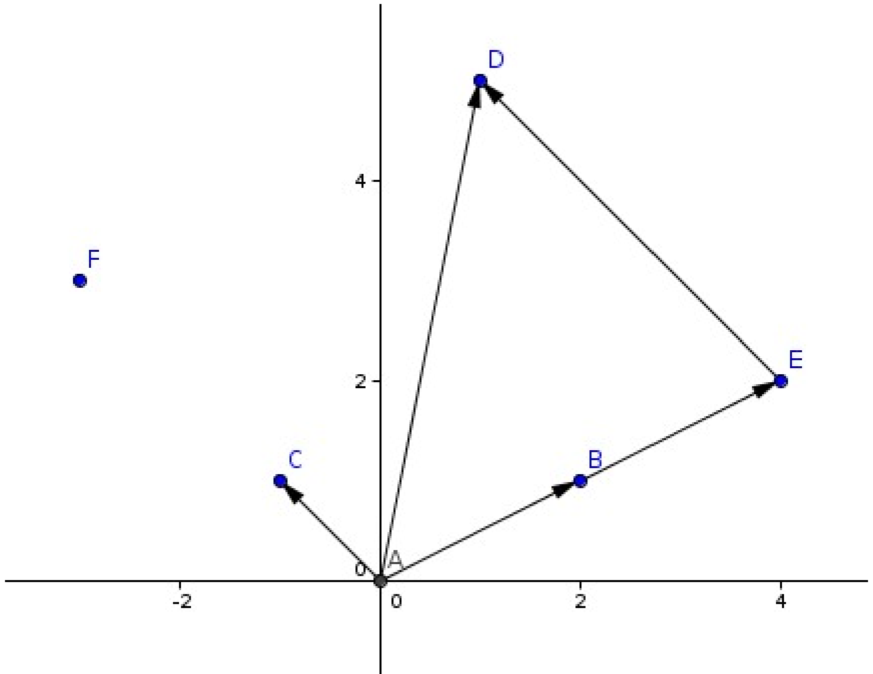


Column and Row Pictures

- Line equation and intersections of lines (Row picture) (2,3)
- Column picture

$$\begin{aligned} 2x - y &= 1 \\ x + y &= 5 \end{aligned}$$





Elimination

- We are typically interested in large systems of equations (2 equations go to "n") (u, v, w) = (1, 1, 2)

$$\begin{aligned} 2u + v + w &= 5 \\ 4u - 6v + 0w &= -2 \\ -2u + 7v + 2w &= 9 \end{aligned}$$

- Solution is obtained by elimination instead of explicitly finding the inverse using Cramer's rule
- Elimination results in
  - pivots (first non-zero in each row)
  - Requires multipliers (entry to eliminate ÷ pivot)
  - $Ax = b$ changes to $Ux = c$

Elimination

- Elimination may lead to permanent failure: No solutions

$$\begin{aligned} x - 2y &= 1 \\ 3x - 6y &= 10 \end{aligned}$$

- Elimination may lead to failure: infinite solutions

$$\begin{aligned} x - 2y &= 1 \\ 3x - 6y &= 3 \end{aligned}$$

- Elimination may lead to *temporary* failure: one solution (3,2)

$$\begin{aligned} 0 + 2y &= 4 \\ 3x - 2y &= 5 \end{aligned}$$

Elimination

- Ax = b becomes upper triangular Ux = c after elimination from top to bottom
- Row(i) = Row(i) - L_{ij} times Row(j) to make (i,j) entry zero (note the *subtract*). Also, augment: [A | b]
- L_{ij} = entry to eliminate in row(i,j) ÷ pivot
  - Pivots cannot be zero
  - A zero in the pivot can be repaired if there is a non-zero below by interchange
- Ux = c is solved bottom to top
- When breakdown is permanent, we have zero solutions or infinite solutions

Elimination

- Please solve
- (u, v, w) = (3, 1, 0)

$$
\begin{aligned}
2u - 3v + 0w &= 3 \\
4u - 5v + w &= 7 \\
2u - v - 3w &= 5
\end{aligned}
$$

Elimination Matrices

- Start with the identity matrix E_{ij}= eye(3)

$$\begin{aligned} 2u - 3v + 0w &= 3 \\ 4u - 5v + w &= 7 \\ 2u - v - 3w &= 5 \end{aligned}$$

- Put $-L_{ij}$ in the location (i,j) (note the *minus* sign)
- For a 3x3 matrix, elimination results in three matrices E_{21}, E_{31}, E_{32} that are multiplied in order to produce U: upper triangular: EA = U
  - $E = E_{32} * E_{31} * E_{21}$

Elimination

- With no shuffling of rows
  - $Ax = b$ changes to $EAx = Ux = c$, via E matrices which are lower triangular
  - $EA = U$ implies $A = E^{-1} U$
  - Amazingly the inverse of $E = L$ can be written down simply by inspection (it was the multipliers, who would have guessed!); L is also lower triangle
- To solve $Ax = b$
  - Find L and U (approximately n^3 computations)
  - $Lc = b$, and then $Ux = c$ (n^2 computations)

Elimination Matrices

- Elimination is also the way to find inverse
- $AA^{-1} = A[x1\ x2\ x3] = [Ax1\ Ax2\ Ax3] = [e1\ e2\ e3]$
- Solve three problems with LU decomposition?
- Gauss-Jordan solves all n equations together
  - Go top to bottom to produce U
  - (Jordan) Go bottom to top to produce zeros above the pivots
  - (Jordan) Divide each row by its pivot
  - Works because $[A\ I]$ goes to $A^{-1}[A\ I] = [I\ A^{-1}]$

- How to tell if ATA is invertible in every single case?
- Notion of vector space and subspace
- A matrix induces subspaces
    - column space and null space

Vector Spaces

- A vector space has eight rules
  - +: Commutativty, associativity, 0, inverse
  - Scalar *: Associativity, Distributivity over the addition of scalars, Distributity over addition of vectors, Unique 1
- Most common R^n
  - but also consider M: the vector space of all real 2 by 2 matrices, or F: all real functions
- Subspace: cv+dw also is in the subspace
  - The combinations of the columns of A form the column space C(A) which is a subspace of R^m

Subspaces

- Consider matrices from:
  - (u+2v, 3u+6v): We are looking at the line u+2v=0 (this is a subspace of R^2)
  - (x+2y+3z): We are looking at the plane in R^3
  - (u+2v, 3u+8v, 2u+4v, 6u+16v)
  - (u+2v+2w+4x, 3u+8v+6w+16x)
  - (u+v+2w+3x, 2u+2v+8w+10x, 3u+3v+10w+13x)

Elimination

- Elimination efficiently discovers subspace
  - Lets us check if b is in the column space of A
  - Also lets us discover the nullspace of A. We are looking at $Ax = \mathbf{0}$ and looking for nonzero x.

Elimination

- Ax = 0 moves to Ux = 0 moves to Rx = 0

$$\begin{bmatrix} \mathbf{p} & a & b & c & d & e & f \\ 0 & \mathbf{p} & g & h & i & j & k \\ 0 & 0 & 0 & 0 & 0 & \mathbf{p} & l \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

- U is no more upper triangular, but is a staircase matrix called the Echelon matrix

$$\begin{bmatrix} \mathbf{1} & 0 & b' & c' & d' & 0 & f' \\ 0 & \mathbf{1} & g' & h' & i' & 0 & k' \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} & l' \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

- In R, there are zeros above the pivots (which are now 1) and EA = R

- Special solutions read off from R (e.g. [-b’, -g’, 1, 0, 0, 0, 0]^T) are members of null space

Null Space

- The nullspace N(A) is a subspace of R^n. It contains **all** solutions to $Ax = 0$
  - If $n>m$, we are sure to have a nonzero x
- Elimination produces echelon matrix U, and a row reduced R with pivot columns and free columns
  - Every free column of U or R leads to a special solution
  - The complete solution to $Ax=\mathbf{0}$ is a combination of these special columns
  - The number of free columns is termed nullity

Necessary Definitions

- A sequence of vectors v_i is *linearly independent* if the only combination that gives the zero vector is $0v_1 + 0v_2 + \dots$
  - The pivot columns are linearly independent
  - The special solutions are linearly independent
- A set of vectors *spans* a space if their linear combination fills the space
  - A *basis* for a vector space is a sequence of linearly independent vectors that spans the space
  - The *dimension* of a space is the number of vectors in every basis

Rank

- Each of these two spaces are discovered by elimination, but we move one step further
  - $Ax = 0$ moves to $Ux = 0$ moves to $Rx = 0$
  - r pivots, r independent rows, r independent columns
- The "true size" of A is r!
  - The rank of the matrix is the number of pivots
  - The free columns are linear combinations of the pivot columns and can be "read off" from R (that's why we did the special solutions)
  - These numbers add up beautifully: Rank (r) + nullity (n – r) = n

Particular Solution

- Consider
  - (x+y=b1, x+2y=b2, -2x + -3y = b3)
  - (x+3y+2w=1, z+4w=6, x+3y+z+6w=7)
  - (x + y +z = 3, x+2y-z =4)
- We proceed with elimination and rref as usual but with an extended column leading to Rx=d
- The particular solution sets free variables to zero, and pivot variables from d
  - Zero rows from R must be zero in d

Procedure

- Ax=b moves to Ux=c moves to Rx=d
- A is appended by the column vector, and the same row reduction is performed (rref)
- Solution only if $c_4=0$

$$\begin{bmatrix} \mathbf{p} & a & b & c & d & e & f & c_1 \\ 0 & \mathbf{p} & g & h & i & j & k & c_2 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{p} & l & c_3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_4 \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{1} & 0 & b' & c' & d' & 0 & f' & d_1 \\ 0 & \mathbf{1} & g' & h' & i' & 0 & k' & d_2 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} & l' & d_3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

- The particular solution can be read off from d
  - (e.g. $[d_1, d_2, 0, 0, 0, d_3, 0]^T$)
- Complete solution sum of the particular

Summary

- With full column rank (r=n), all columns are pivot columns, null space is 0, if there is a solution, there will be exactly one.
- With full row rank (r=m), all rows are pivot rows, and $C(A) = R^m$
- With $r < m$ and $r < n$, Ax=b has 0 solutions or infinite solutions
- With $r = m$ and $r = n$, Ax=b has 1 solution

Particular Solution

- Consider
  - (x+y= b1, x+2y= b2, -2x + -3y = b3)
  - (x+3y+2w= 1, z+4w= 6, x+3y+z+6w= 7)
  - (x + y +z= 3, x+2y-z= 4)
- We proceed with elimination and rref as usual but with an extended column leading to Rx=d
- The particular solution sets free variables to zero, and pivot variables from d
  - Zero rows from R must be zero in d

Summary

- With full column rank (r=n), all columns are pivot columns, null space is 0, if there is a solution, there will be exactly one.
- With full row rank (r=m), all rows are pivot rows, and $C(A) = R^m$
- With $r < m$ and $r < n$, Ax=b has 0 solutions or infinite solutions
- With $r = m$ and $r = n$, Ax=b has 1 solution

Procedure

- Ax=b moves to Ux=c moves to Rx=d
- A is appended by the column vector, and the same row reduction is performed (rref)
- Solution only if $c_4=0$

$$\begin{bmatrix} \mathbf{p} & a & b & c & d & e & f & c_1 \\ 0 & \mathbf{p} & g & h & i & j & k & c_2 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{p} & l & c_3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_4 \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{1} & 0 & b' & c' & d' & 0 & f' & d_1 \\ 0 & \mathbf{1} & g' & h' & i' & 0 & k' & d_2 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} & l' & d_3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

- The particular solution can be read off from d
  - (e.g. $[d_1, d_2, 0, 0, 0, d_3, 0]^T$)
- Complete solution = sum of the particular

Motivation: A^T

- The rank + nullity theorem has n in it, not m. What happened to m?
- Consider $x - 3y - 4z = 0$: What is the nullspace? What is the basis? Is $[6\ 4\ 5]^T$ in the null space?
  - If not, then where is it?
  - Punch line: Every vector x in R^n is made of a combination of a part in the null space and a part in the column space of A^T (this is the row space!)

Best x when Ax ≠ b

- With n linearly independent vectors (a_1, a_2, ... a_n) in R^m, find the combination $p = w1a_1 + ... + wna_n$ closest to a given vector b
- Solution: The error vector b – Aw is perpendicular to the subspace, and thus to each of the n vectors a_i
  - Write $ai^T (b - Aw) = 0$
  - The matrix A^TA is invertible! (null space of A^TA is the same as that of A

Definitions

- Two vectors are orthogonal if $v^T w = 0$
- Two subspaces V and W of a vector space are orthogonal if every vector v in V is orthogonal to every vector w in W
- The orthogonal complement (“$V^{\perp}$”)of a subspace V contains every vector that is perpendicular to V (pronounced V-perp)

Recap

- If a matrix has **full row rank** (r=m)
  - All rows of A are pivot rows (no zero rows)
  - There are n-r special solutions
  - Ax = b always has ∞ solutions

Recap

- Four fundamental subspaces
  - And their orthogonality
  - And their dimensions
  - And Rank (r) + nullity (n – r) = n
  - And the basis vectors
- Reduction of Ax = b to Rx = d
  - Nullspace matrix
  - Reading of all solutions

Recap

- If a matrix has **full row rank** (r=m)
  - All rows of A are pivot rows (no zero rows)
  - There are n-r special solutions
  - Ax = b always has ∞ solutions

Summary

- If a matrix has **full column rank** (r=n)
  - All columns of A are pivot columns
  - There are no free variables or special solutions
  - N(A) contains only the zero vector
  - Ax = b has zero solutions **or** one solution
  - The square matrix $A^T A$ has full rank

Summary

- How many solutions for Ax = b
  - If r = m = n, exactly 1 solution (square matrix)
  - If r = m, m < n, ∞ solutions (short and fat)
  - If r = n, n < m, 0 or 1 solution (tall and thin)
  - If r < m, r < n, 0 or ∞ solutions (can shrink)
- Shape

$$R = \begin{bmatrix} I \end{bmatrix} \quad R = \begin{bmatrix} I & F \end{bmatrix} \quad R = \begin{bmatrix} I \\ 0 \end{bmatrix} \quad R = \begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix}$$

Definitions

- Two vectors are orthogonal if $v^T w = 0$
- Two subspaces V and W of a vector space are orthogonal if every vector v in V is orthogonal to every vector w in W
  - Floor, and line of intersection of two walls
- The orthogonal complement (“$V^{\perp}$”)of a subspace V contains every vector that is perpendicular to V (pronounced V-perp)

FTLA-2

- Null space is the **orthogonal complement** of the row space.
  - **If** *x* perpendicular to the rows, $Ax = 0$
  - **If v is orthogonal to the nullspace, it must be in the row space** (otherwise we could add v as an additional row increasing the rank and breaking $r + (n-r) = n$
- $N(A)$ is the orthogonal complement of $C(A^T)$
- $N(A^T)$ is the orthogonal complement of $C(A)$

Today: Agenda

- Summary of complete solution to Ax=b
- What happens when we can't solve Ax = b?

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} x = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$$

Agenda

- Find the "best" line that approximates e^x
- Before that find the best line that approximates the points (-1, 1) (1, 1) (2, 3)
    - In earlier class we had seen that this is the solutions to the normal equation $x = (A^TA)^{-1}A^Tb$ where A = [1 -1; 1 1; 1 2] and b = [1 1 3]'
    - Slope = 4/7, Y intercept = 9/7

Null Space of A^TA

- The null space of A^TA is the same as that of A
  - Thus if A has linearly independent columns, the matrix is invertible
  - This results in the so-called normal equations in linear regression

Projection

- The projection of b onto a subspace is given by $A(A^TA)^{-1}A^Tx$
- If the matrix A contains orthonormal columns, then much of the terms vanish since $Q^TQ=I$

Story So Far

- Row Transformations A -> U -> R = [I F; 0 0]
- Four vector spaces and their basis
  - Row space, col space, null space, left null space
  - Basis: Pivot rows of (R and A), pivot cols of A (not R), special solutions from F and I, last (m-r) rows of the elimination matrix E
  - This leads to the rank nullity theorem r + nullity = n
- Why do we conceptualize such spaces?

Story So Far

- The equation Ax = b is fundamental
- This solution can be "solved": left and right inverse
  - If BA =I then clearly BAx = Bb and so x = Bb. One such B is inv(A'A)A'. But does B always exist? What is B?
  - If AC = I, then define x = Cb, thus Ax = ACb = b. We have a solution. Does C exist? What is C?
- By knowing the concept of the four fundamental spaces we can be sure when B and C exist
- Such B and exist if the ranks are as large as possible

Existence of B and C

- If A has full column rank, then B exists
  - B = inv(A'A)A' and inv(A'A) exist
  - B provides either the unique solution or the best possible solution
- If A has full row rank, then C exists
  - C = A'inv(AA').
  - There are exists at least one solution if b is in the column space
  - Exists infinite solutions and C is the "best" possible solution
- Only a square matrix can have

Inference

- Given a query image (may not be a face), it is projected onto the subspace of linearly independent vectors
  - If the projection is in the null space, then we declare it to be the non-face
  - Otherwise, we find the nearest face (could be an impostor)