Least Squares

Reading: Strang 4.3

Learning objective: Be able to find the least squares solution to a system of linear equations and know when it is appropriate to do so.

Review

If $A\vec{x} = \vec{b}$ does not have a solution, we can seek $\hat{\mathrm{x}}$ to minimize $\|\vec{b} - A\hat{\mathrm{x}}\|$.

Such an $\hat{\mathrm{x}}$ is called a least squares solution to $A\vec{x} = \vec{b}$ because it minimizes the sum of the squares of the components of the error vector $\vec{e} = \vec{b} - A\hat{\mathrm{x}}$.

$A\hat{\mathrm{x}}$ is the closest point in the column space to $\vec{b}$.

It is the projection of $\vec{b}$ onto $C(A)$.

Review

$A\hat{\mathbf{x}}$ is the projection of $\vec{b}$ onto $C(A)$.

Thus the error vector $\vec{b} - A\hat{\mathbf{x}}$ will be orthogonal to the column space. It will be in the left nullspace.

$$A^T(\vec{b} - A\hat{\mathbf{x}}) = \vec{0} \implies A^T A\hat{\mathbf{x}} = A^T\vec{b}$$

We can find the least squares solution by solving this normal equation.

Shortcut: when $A\vec{x} = \vec{b}$ doesn't have a solution, multiply both sides by A^T and solve the resulting equation.

Projection of $\vec{b}$ onto line through $\vec{a}$		Projection of $\vec{b}$ onto column space of A
$\hat{x} \cdot \vec{a}$	point in the subspace	$A\hat{\mathbf{x}}$
$\vec{a}^T(\vec{b} - \hat{x} \cdot \vec{a}) = 0$	difference is ortho. to subspace	$A^T(\vec{b} - A\hat{\mathbf{x}}) = \vec{0}_n$
$\hat{x} = \dfrac{\vec{a}^T\vec{b}}{\vec{a}^T\vec{a}}$	solving for $\hat{x}$	$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\vec{b}$
$\vec{a} \cdot \dfrac{\vec{a}^T\vec{b}}{\vec{a}^T\vec{a}}$	projection	$A(A^TA)^{-1}A^T\vec{b}$

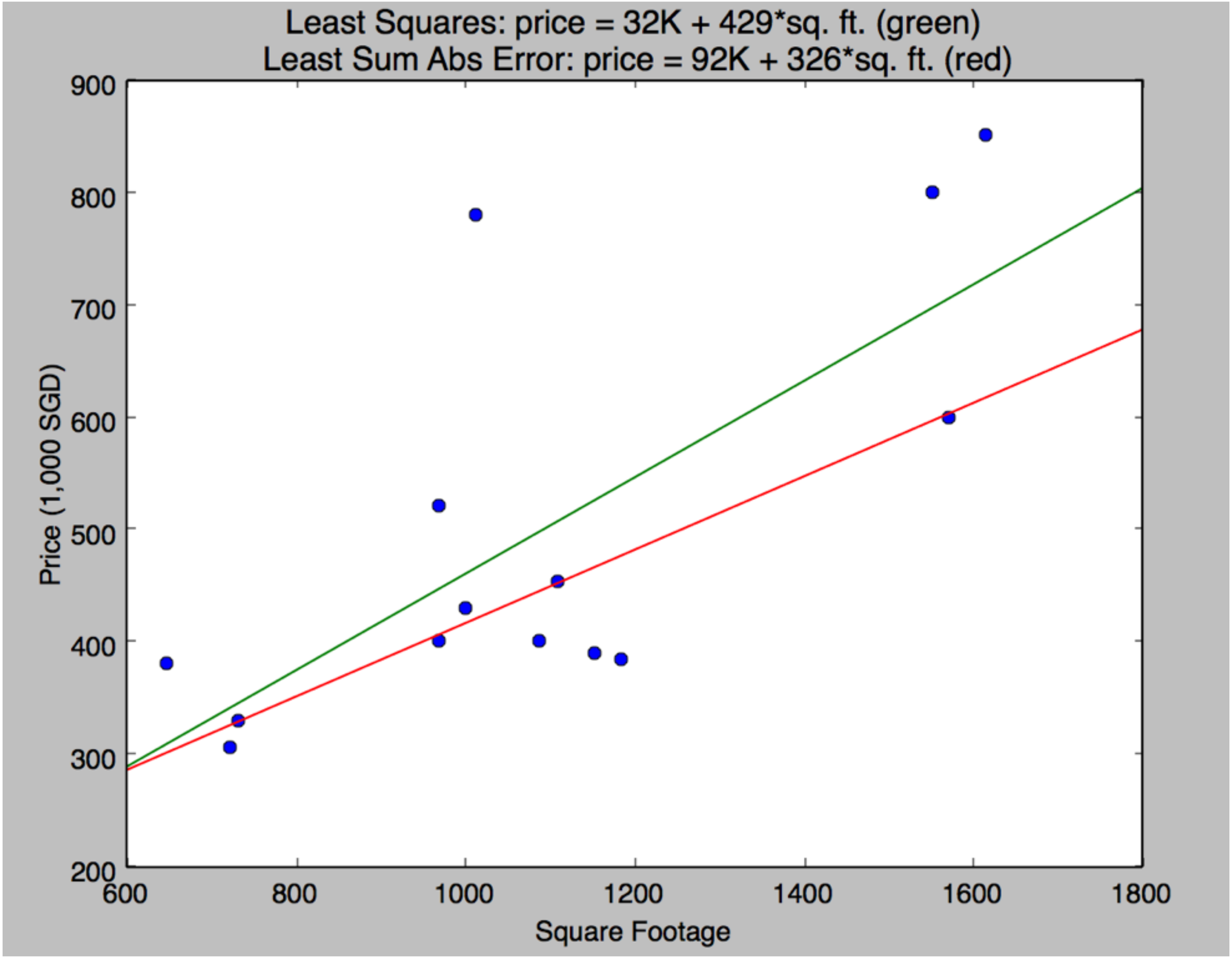
Question

In least squares we look for the $\hat{\mathrm{x}}$ which minimizes $\|\vec{b} - A\hat{\mathrm{x}}\|^2$, the sum of the squares of the errors.

Why don't we instead look for the solution which minimizes the sum of the absolute values of the errors?

Answer: This is also a perfectly reasonable thing to do, and more appropriate for some applications.

However, the solution which minimizes the sum of the absolute values of the errors is more difficult to compute.


Least Squares: price = 32K + 429*sq. ft. (green)
Least Sum Abs Error: price = 92K + 326*sq. ft. (red)
Price (1,000 SGD)
Square Footage
900
800
700
600
500
400
300
200
600
800
1000
1200
1400
1600
1800

Today

We will wrap up our discussion of least squares with two further points.

§ Discussion of the assumption that $A^T A$ is invertible.

§ Application of least squares to data fitting.

Invertibility Assumption

$$A^T A\hat{\mathbf{x}} = A^T\vec{b}$$

The normal equation always has a solution.

When $A^T A$ is invertible, the normal equation has a unique solution.

$$\hat{\mathbf{x}} = (A^T A)^{-1}A^T\vec{b}$$

Note: Even when $A^T A$ is invertible it is faster to solve the normal equations by Gaussian elimination!

Projection onto Column Space

When $A^T A$ is invertible, we can write down the projection matrix onto the column space of A.

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \vec{b}$$

The closest point in the column space to $\vec{b}$ is then

$$A\hat{\mathbf{x}} = A(A^T A)^{-1} A^T \vec{b}$$

The projection matrix P onto $C(A)$ is

$$P = A(A^T A)^{-1} A^T$$

Projection onto Column Space

The projection matrix P onto $C(A)$ is

$$P = A(A^TA)^{-1}A^T$$

Note that

$$\begin{aligned} P^2 &= A(A^TA)^{-1}A^TA(A^TA)^{-1}A^T \\ &= A(A^TA)^{-1}A^T \\ &= P \end{aligned}$$

Also P is symmetric.

Invertibility Assumption

When $A^T A$ is invertible, the normal equation has a unique solution.

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \vec{b}$$

As $A^T A$ is always square, it will be invertible iff

$$N(A^T A) = \{\vec{0_n}\}$$

Theorem: $N(A^T A) = N(A)$

On the problem set, you give one proof of this theorem. I'll give another one here.

Theorem: $N(A^T A) = N(A)$

Proof:

$N(A) \subseteq N(A^T A)$

If $A\vec{u} = \vec{0}_m$ then $A^T A\vec{u} = A^T\vec{0}_m = \vec{0}_n$.

$N(A^T A) \subseteq N(A)$

If $A^T A\vec{u} = \vec{0}_n$ then $\vec{u}^T A^T A\vec{u} = 0$.

$\vec{u}^T A^T A\vec{u} = \|A\vec{u}\|^2$ is only zero when $A\vec{u} = \vec{0}_m$.

Invertibility Assumption

As $A^T A$ is always square, it will be invertible iff

$$N(A^T A) = \{\vec{0}_n\}$$

Theorem: $N(A^T A) = N(A)$

Thus $A^T A$ is invertible iff $N(A) = \{\vec{0}_n\}$ that is, if the columns of A are linearly independent.

Invertibility Assumption

Thus A^TA is invertible iff $N(A) = \{\vec{0}_n\}$ that is, if the columns of A are linearly independent.

§ In practice this will almost always be the case. Usually A is a tall skinny matrix.

$$\begin{bmatrix} 1 & 1614 \\ 1 & 968 \\ 1 & 1184 \\ 1 & 968 \\ 1 & 1000 \\ 1 & 1152 \\ 1 & 1087 \\ 1 & 1108 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 850 \\ 400 \\ 385 \\ 520 \\ 430 \\ 390 \\ 400 \\ 453 \end{bmatrix}$$

housing prices example

Invertibility Assumption

Thus A^TA is invertible iff $N(A) = \{\vec{0}_n\}$ that is, if the columns of A are linearly independent.

§ If the columns of A are linearly dependent, we can remove redundant columns. This does not change the column space.

Doing this, the projection of $\vec{b}$ onto the column space will not change.

Application of Least Squares

Reading: Strang 4.3

Learning objective: Be able to apply the least squares method to data fitting problems.

Example: Papaya Tree Height

Here is some data I made up about the growth of a papaya tree.

time (months)	height (cm)
1	5
2	11
3	16
4	19

What line of the form

$$\text{height} = a_0 + a_1 \cdot \text{time}$$

best fits the data in the least squares sense?

time (months)	height (cm)
1	5
2	11
3	16
4	19

What line of the form

$$\text{height} = a_0 + a_1 \cdot \text{time}$$

best fits the data in the least squares sense?

The question asks for the least squares solution of

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \\ 16 \\ 19 \end{bmatrix}$$

The question asks for the least squares solution of

$$\begin{bmatrix}1 & 1\\1 & 2\\1 & 3\\1 & 4\end{bmatrix}\begin{bmatrix}a_0\\a_1\end{bmatrix}=\begin{bmatrix}5\\11\\16\\19\end{bmatrix}$$

To do this we solve the normal equation

$$A^TA\vec{x}=A^T\vec{b}$$

$$A^TA=\begin{bmatrix}1 & 1 & 1 & 1\\1 & 2 & 3 & 4\end{bmatrix}\begin{bmatrix}1 & 1\\1 & 2\\1 & 3\\1 & 4\end{bmatrix}=\begin{bmatrix}4 & 10\\10 & 30\end{bmatrix}$$

$$A^T\vec{b}=\begin{bmatrix}1 & 1 & 1 & 1\\1 & 2 & 3 & 4\end{bmatrix}\begin{bmatrix}5\\11\\16\\19\end{bmatrix}=\begin{bmatrix}51\\151\end{bmatrix}$$

To do this we solve the normal equation

$$A^T A\vec{x} = A^T\vec{b}$$

$$\begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 51 \\ 151 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 10 & \mathbf{51} \\ 10 & 30 & \mathbf{151} \end{bmatrix} \xrightarrow[R_2' = R_2 - \frac{5}{2}R_1]{} \begin{bmatrix} 4 & 10 & \mathbf{51} \\ 0 & 5 & \mathbf{23.5} \end{bmatrix}$$

Solving by back substitution:

$$a_2 = \frac{47}{10} \qquad a_1 = 1$$

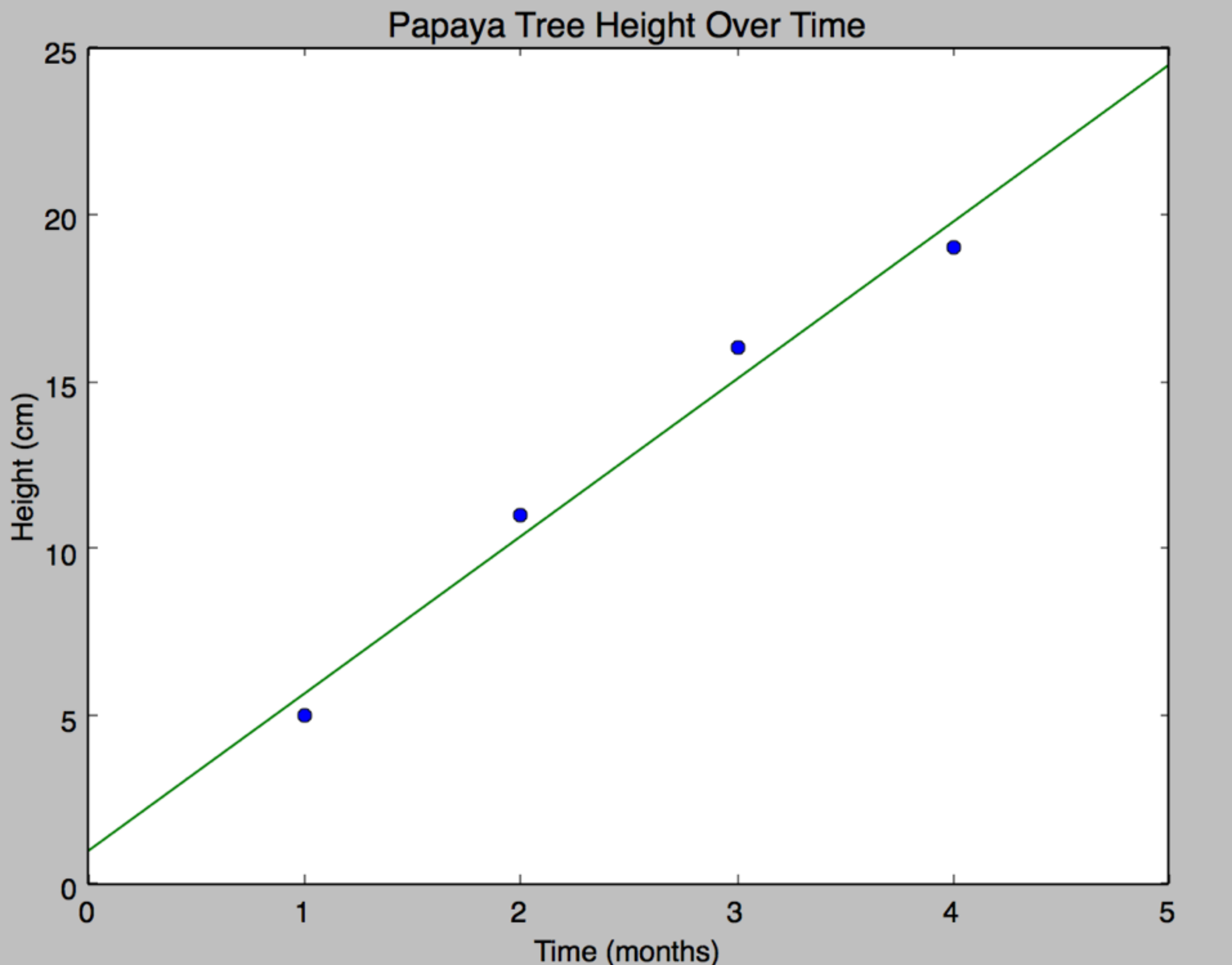
Now we have found $\hat{\mathbf{x}} = (1, \frac{47}{10})$.

The projection of $(5, 11, 16, 19)$ onto the column space of

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

is

$$A\hat{\mathbf{x}} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ \frac{47}{10} \end{bmatrix} = \begin{bmatrix} 5.7 \\ 10.4 \\ 15.1 \\ 19.8 \end{bmatrix}$$


Papaya Tree Height Over Time
Height (cm)
Time (months)
0
5
10
15
20
25
0
1
2
3
4
5

Prediction

From the historical data, we have found the line of best fit

$$\text{height} = 1 + 4.7 \cdot \text{time}$$

Now we can use this formula to predict the height of the papaya tree in the future!

After one year, we predict the papaya tree will be $1 + 4.7 \cdot 12 = 57.4$ cm tall.

This is a very typical application of least squares.

Housing Prices

The exact same procedure can work with more independent variables.

Let's expand our housing prices example to include more variables and look for a formula

$$\text{price} = a_0 + a_1 \cdot \text{sq. foot} + a_2 \cdot \text{beds} + a_3 \cdot \text{baths} + a_4 \cdot \text{location}$$

Here location takes on the value 0 or 1 based on my opinion of the centrality of the location.

Example:

Woodlands = 0

Toa Payoh = 1

$$
\begin{bmatrix}
1 & 1076 & 3 & 2 & 0 \\
1 & 1302 & 3 & 2 & 0 \\
1 & 1270 & 3 & 2 & 1 \\
1 & 731 & 2 & 2 & 1 \\
1 & 1571 & 3 & 3 & 0 \\
1 & 699 & 1 & 1 & 1 \\
1 & 484 & 1 & 1 & 1 \\
1 & 1367 & 4 & 2 & 0
\end{bmatrix}
\begin{bmatrix}
a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4
\end{bmatrix}
=
\begin{bmatrix}
3.65 \\ 3.95 \\ 5.99 \\ 5.5 \\ 6.2 \\ 2.9 \\ 2.88 \\ 8.0
\end{bmatrix}
$$

sq. ft.

beds

baths

location

price
100,000 SGD

We look for the least squares solution to this equation.

Best Fit

The formula that best fits the data is

$$\text{price} = -1.94 + 0 \cdot \text{sq. ft} + 1.7 \cdot \text{beds} + .91 \cdot \text{baths} + 2.05 \cdot \text{location}$$

$$\begin{bmatrix} 3.65 \\ 3.95 \\ 5.99 \\ 5.5 \\ 6.2 \\ 2.9 \\ 2.88 \\ 8.0 \end{bmatrix} \qquad \begin{bmatrix} 4.87 \\ 4.78 \\ 6.84 \\ 5.25 \\ 5.6 \\ 2.54 \\ 2.63 \\ 6.55 \end{bmatrix}$$

actual prices　　　　projected prices

Final Review

Major Topics

§ Vectors

Addition, scalar multiplication, taking linear comb.

Dot Product

Geometric view of sets of vectors

§ Solving systems of linear equations

Gaussian elimination

Determining the number of solutions

LU Decomposition

Major Topics

§ Matrix Multiplication

Column picture, Row picture

Row-Column picture (dot products)

Column-Row picture (outer products)

§ Invertibility

Finding the inverse (Gauss-Jordan Elimination)

Conditions Equivalent to Invertibility

Major Topics

- § Determinants
  - Defining Properties
  - Computing by Gaussian Elimination
  - Cofactor formula
- § Vector Spaces
  - Recognizing a Subspace
  - Determining linear independence
  - Finding a basis for a vector space
  - Dimension Theorem

Major Topics

§ 4 Subspaces associated with a matrix

- How to find a basis and compute their dimension
- Rank revealing factorization
- Orthogonality relations

§ Orthogonality

- Orthogonal Complement
- Projection onto a line/subspace
- Least squares solution

Systems of Linear Equations

To solve a system of linear equations $A\vec{x} = \vec{b}$ we can do Gaussian elimination on the augmented matrix

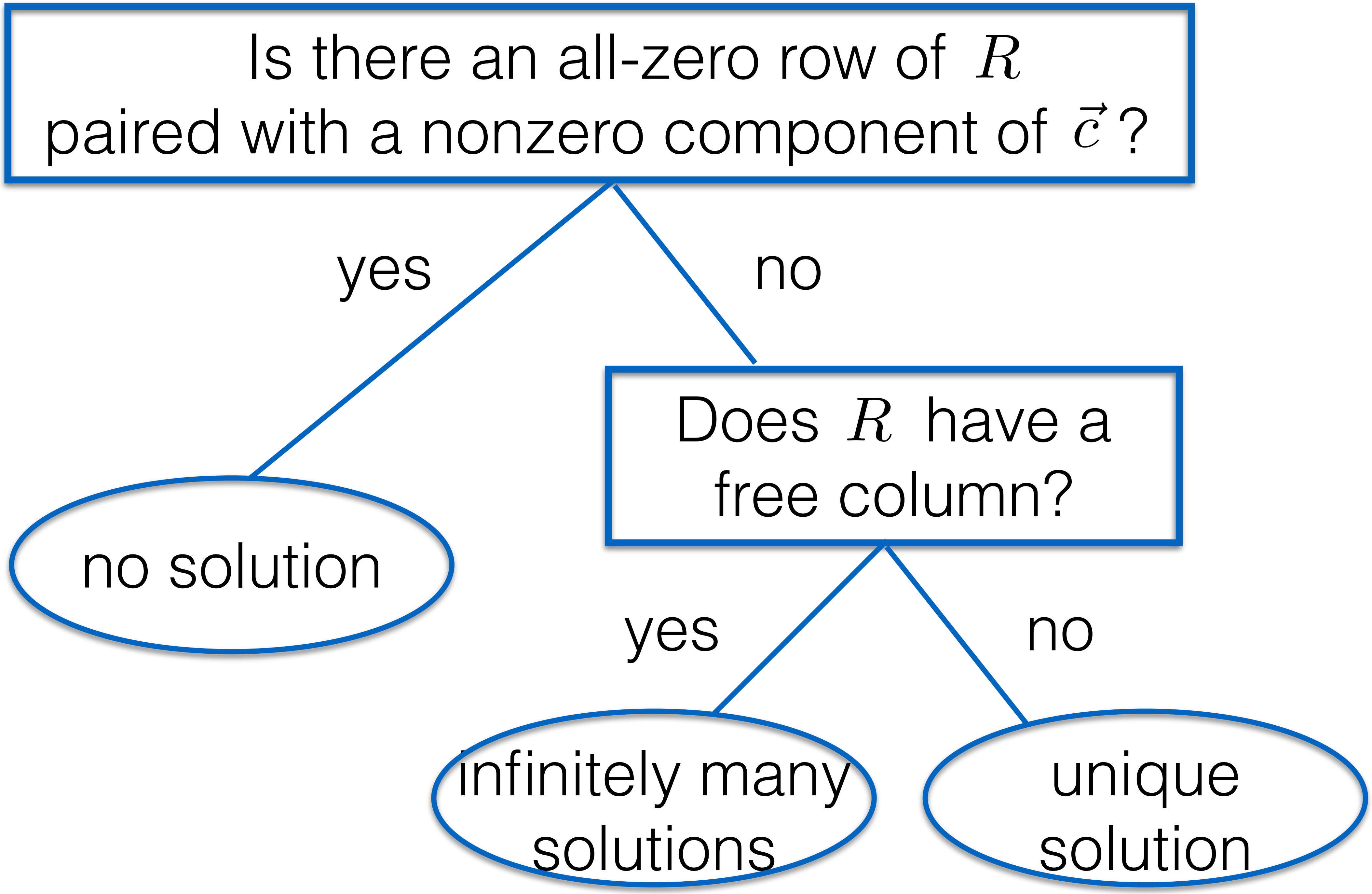
$$[A \mid \vec{b}] \xrightarrow[\text{G.E.}]{} [R \mid \vec{c}]$$

row echelon form

Don't forget to also modify $\vec{b}$ by the row operations.

The solution set to $A\vec{x} = \vec{b}$ is the same as the solution set to $R\vec{x} = \vec{c}$.

Say that R is in row echelon form. Here is a flowchart to determine the number of solutions to $R\vec{x} = \vec{c}$.



$$[A \mid \vec{b}] \xrightarrow{\text{G.E.}} [R \mid \vec{c}]$$

row echelon form

$\boxed{\text{Is there an all-zero row of } R \text{ paired with a nonzero component of } \vec{c}\,?}$

This is the only impediment to $A\vec{x} = \vec{b}$ having a solution.

Here is a fancy way of saying this (Fredholm's Alternative):

Exactly one of these problems has a solution

$$A\vec{x} = \vec{b} \qquad \text{OR} \qquad \vec{y}^T A = \vec{0}_n^T \text{ with } \vec{y}^T\vec{b} \neq 0$$

Gaussian elimination constructs such a $\vec{y}$ when $A\vec{x} = \vec{b}$ has no solution.

Fredholm's Alternative

Exactly one of these problems has a solution

$$A\vec{x} = \vec{b} \qquad \text{OR} \qquad \vec{y}^T A = \vec{0}_n^T \text{ with } \vec{y}^T\vec{b} \neq 0$$

The orthogonal complement of the left nullspace of A is the column space of A .

If $\vec{b} \notin C(A)$ then it cannot be orthogonal to the left nullspace .

Thus if $\vec{b} \notin C(A)$ there exists $\vec{y} \in N(A^T)$ with $\vec{y}^T\vec{b} \neq 0$.

Systems of Linear Equations

We can apply some of our new words to systems of linear equations.

Let A be an m-by-n matrix of rank r.

If $r = m$ then $A\vec{x} = \vec{b}$ will always have a solution.

$$C(A) = \mathbb{R}^m$$

If $r = n$ then $A\vec{x} = \vec{b}$ will either have no solution or a unique solution.

$$N(A) = \{\vec{0}_n\}$$

Four Subspaces

Let A be an m-by-n matrix with r pivots.

	$C(A^T)$	$N(A)$	$C(A)$	$N(A^T)$
space	$\mathbb{R}^n$	$\mathbb{R}^n$	$\mathbb{R}^m$	$\mathbb{R}^m$
row ops preserve	yes	yes	no	no
dimension	r	$n-r$	r	$m-r$

The row space and nullspace are orthogonal complements.

The column space and left nullspace are orthogonal complements.

Four Subspaces

Finding a basis for the four subspaces of a matrix A with a row echelon form R.

Row space: A basis is given by the nonzero rows of R.

Column space: A basis is given by the columns of A corresponding to pivot columns of R.

Nullspace: A basis is given by the special solutions to

$$R\vec{x} = \vec{0}$$

Left nullspace: If $EA = R$ then a basis is given by the rows of E corresponding to zero rows of R.

Conditions Equivalent to Invertibility

Let's revisit our "big list" of conditions equivalent to invertibility.

This will take us through a lot of the concepts and vocabulary we have learned over this course.

Invertible: A square matrix A is invertible iff there exists a matrix B such that $AB = I$ and $BA = I$.

Singular: A square matrix that is not invertible.

The Big List

Let A be an n-by-n matrix. The following are equivalent:

- A is invertible.
- Gaussian elimination produces n pivots.
- $A\vec{x} = \vec{0}$ has a unique solution.
- A has a left inverse.
- A has a right inverse.
- The reduced row echelon form of A is the identity matrix.
- A is the product of elementatry matrices.
- A^T is invertible.
- $\det(A) \neq 0$.

The Big List

Let A be an n-by-n matrix. The following are equivalent:

- A is invertible.
- The columns of A are linearly independent .
- $N(A) = \{\vec{0}_n\}$.
- The rows of A are linearly independent .
- $N(A^T) = \{\vec{0}_n\}$.
- $A\vec{x} = \vec{b}$ has a solution for every $\vec{b} \in \mathbb{R}^n$.
- $C(A) = \mathbb{R}^n$.
- $\mathrm{rank}(A) = n$.
- The orthogonal complement of the row space is $\{\vec{0}_n\}$.

Inverse Problem

Here is an interesting problem from the discussion forum.

Let A be an invertible 2n-by-2n matrix of the following form:

$$A = \begin{bmatrix} B & C \\ D & 0_{n\times n} \end{bmatrix}$$

Show that A^{-1} is of the following form:

$$A^{-1} = \begin{bmatrix} 0_{n\times n} & E \\ F & G \end{bmatrix}$$

Solution 1

$$A = \begin{bmatrix} B & C \\ D & 0_{n \times n} \end{bmatrix}$$

We are given that A is invertible, this means the columns of A are linearly independent.

In particular the n many columns of C are linearly independent $\Longrightarrow$ the columns of C span $\mathbb{R}^n$.

Any vector whose last n coordinates are zero can be written as a lin. comb. of columns from the right half of A.

Solution 1

$$A = \begin{bmatrix} B & C \\ D & 0_{n\times n} \end{bmatrix}$$

Any vector whose last n coordinates are zero can be written as a lin. comb. of columns from the right half of A.

If the last n coordinates of $\vec{b}$ are zero there is a solution to $A\vec{x} = \vec{b}$ where the first n coordinates of $\vec{x}$ are zero.

Solution 2

Let's first imagine that A is 2-by-2.

$$A = \begin{bmatrix} b & c \\ d & 0 \end{bmatrix}$$

If A is invertible then $bc \neq 0$.

$$A^{-1} = \begin{bmatrix} 0 & d^{-1} \\ c^{-1} & -c^{-1}bd^{-1} \end{bmatrix}$$

Solution 2

When two 2n-by-2n matrices are partitioned into n-by-n blocks, we can multiply them as if each block was just a number.

$$\begin{bmatrix} P & Q \\ R & S \end{bmatrix} \begin{bmatrix} W & X \\ Y & Z \end{bmatrix} = \begin{bmatrix} PW + QY & PX + QZ \\ RW + SY & RX + SZ \end{bmatrix}$$

Thus

$$A = \begin{bmatrix} B & C \\ D & 0_{n\times n} \end{bmatrix} \qquad A^{-1} = \begin{bmatrix} 0_{n\times n} & D^{-1} \\ C^{-1} & -C^{-1}BD^{-1} \end{bmatrix}$$

It is left to see why C and D are invertible when A is.

Solution 2

$$A = \begin{bmatrix} B & C \\ D & 0_{n\times n} \end{bmatrix}$$

It is left to see why C and D are invertible when A is.

We have already seen that the columns of C are linearly independent, thus C is invertible.

$$\begin{bmatrix} B & C \\ D & 0_{n\times n} \end{bmatrix} \begin{bmatrix} I_{n\times n} & 0_{n\times n} \\ -C^{-1}B & I_{n\times n} \end{bmatrix} = \begin{bmatrix} 0_{n\times n} & C \\ D & 0_{n\times n} \end{bmatrix}$$

The last matrix is invertible, thus so is D.

Vector Space Example

The trace of a matrix A, written $\mathrm{Tr}(A)$, is the sum of the diagonal entries of A.

$$\mathrm{Tr}\left(\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\right) = a_{11} + a_{22} + a_{33}$$

Let $M_{3,3}$ be the vector space of 3-by-3 matrices and consider the set

$$S = \{A \in M_{3,3} : \mathrm{Tr}(A) = 0\}$$

Show that S is a subspace and find a basis for S.

Subspace

We can verify that S satisfies the three properties:

1) Nonempty: The sum of the diagonal entries of the zero matrix $\mathbf{0}_{3\times 3}$ is zero, thus $\mathbf{0}_{3\times 3} \in S$.

2) Closure under addition: Let $A, B \in S$. Then

$$\begin{aligned} \mathrm{Tr}(A+B) &= \mathrm{Tr}\left(\begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} \\ a_{31}+b_{31} & a_{32}+b_{32} & a_{33}+b_{33} \end{bmatrix}\right) \\ &= a_{11}+b_{11}+a_{22}+b_{22}+a_{33}+b_{33} \\ &= \mathrm{Tr}(A)+\mathrm{Tr}(B) \\ &= 0 \end{aligned}$$

Subspace

We can verify that S satisfies the three properties:

3) Closure under scalar multiplication: Let $A \in S$ and $c \in \mathbb{R}$

$$\begin{aligned}\mathrm{Tr}(c \cdot A) &= \mathrm{Tr}\left(\begin{bmatrix} ca_{11} & ca_{12} & ca_{13} \\ ca_{21} & ca_{22} & ca_{23} \\ ca_{31} & ca_{32} & ca_{33} \end{bmatrix}\right) \\ &= c \cdot (a_{11} + a_{22} + a_{33}) \\ &= 0\end{aligned}$$

Finding a Basis

Consider a general 3-by-3 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

If $\mathrm{Tr}(A) = 0$ then $a_{11} + a_{22} + a_{33} = 0$.

Let us focus on this linear equation:

$$a_{11} + a_{22} + a_{33} = 0$$

This is like a nullspace problem with two free columns:

a_{22}, a_{33} can be any real number but then $a_{11} = -a_{22} - a_{33}$.

Finding a Basis

Thus a general matrix with trace zero is of the form:

$$\begin{bmatrix} -a_{22}-a_{33} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Let E_{ij} be the matrix which is zero everywhere except for the (i,j) entry, which is one.

By separation of variables, we can write the above matrix as a linear combination of the matrices

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$, and E_{ij} for $i \neq j$.

Spanning

Thus a general matrix with trace zero is of the form:

$$\begin{bmatrix} -a_{22} - a_{33} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

By separation of variables, we can write the above matrix as a linear combination of the matrices

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$, and E_{ij} for $i \neq j$.

Thus the span of these 8 matrices is equal to S .

Linear Independence

We also claim that the 8 matrices

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{and } E_{ij} \text{ for } i \neq j \,.$$

are linearly independent.

We can see this by the easy test.

We also claim that the 8 matrices

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{and } E_{ij} \text{ for } i \neq j .$$

are linearly independent.

The only matrix in this sequence which is nonzero in the (2,2) entry is the matrix

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

Thus its coefficient in a linear combination equal to the zero matrix must be zero.

We also claim that the 8 matrices

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{and } E_{ij} \text{ for } i \neq j \,.$$

are linearly independent.

Similarly, the only matrix in this sequence which is nonzero in the (3,3) entry is the matrix

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

Finally E_{ij} is the only matrix in the sequence which is nonzero in the (i, j) entry for $i \neq j$.

Finding a Basis

Conclusion: The 8 matrices

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{and } E_{ij} \text{ for } i \neq j .$$

are a basis for the subspace $S = \{A \in M_{3,3} : \mathrm{Tr}(A) = 0\}$.

The dimension of S is 8.

Example with proofs

Here is a problem from the 2008-2009 final, Q2:

Show that for any square matrix A, the nullspace of A is contained in the nullspace of A^2.

Proof: If $\vec{x} \in N(A)$ then $A\vec{x} = \vec{0}$. Thus also

$$A^2\vec{x} = A\vec{0} = \vec{0}$$

and $\vec{x} \in N(A^2)$.

Example with proofs

Here is a problem from the 2008-2009 final, Q2:

For a square matrix A, if the column space of A contains a nonzero vector $\vec{x}$ such that $A\vec{x} = \vec{0}$ then

$$\operatorname{rank}(A^2) < \operatorname{rank}(A).$$

Proof: As $\vec{x}$ is in the column space of A,

$$\vec{x} = A\vec{y}$$

for some vector $\vec{y}$.

$$\vec{x} \neq \vec{0} \implies \vec{y} \notin N(A)$$

$$A\vec{x} = \vec{0} \implies \vec{y} \in N(A^2)$$

Example with proofs

Here is a problem from the 2008-2009 final, Q2:

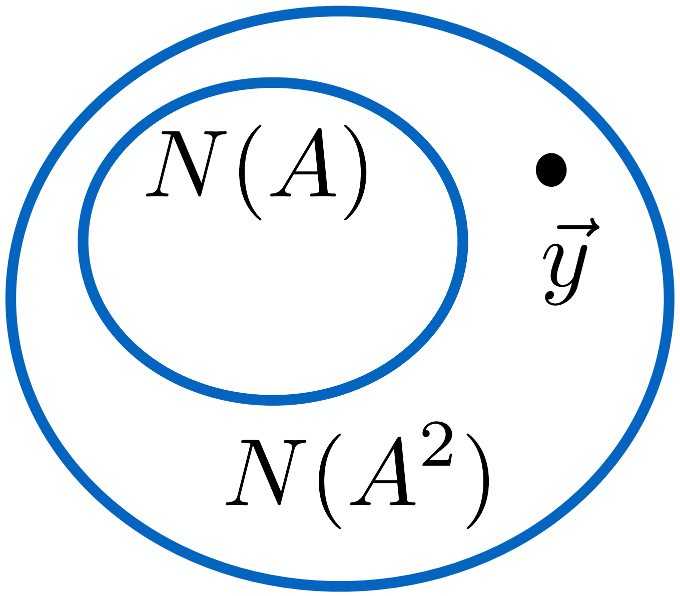
For a square matrix A, if the column space of A contains a nonzero vector $\vec{x}$ such that $A\vec{x} = \vec{0}$ then

$$\text{rank}(A^2) < \text{rank}(A).$$

Proof: From the previous problem $N(A) \subseteq N(A^2)$ thus $N(A) \subsetneq N(A^2)$

This means $\dim(N(A^2)) > \dim(N(A))$ and so $\text{rank}(A^2) < \text{rank}(A).$

$$\text{rank}(B) + \dim(N(B)) = \#\text{ cols of } B$$



The End

Hope you enjoyed the course!