DSC 40A - Extra Practice Session 3

Wednesday, February 2, 2022

Problem 1. Matrix, Vector, Scalar, or Nonsense?

Suppose M is an $m \times n$ matrix, v is a vector in $\mathbb{R}^n$, and s is a scalar. Determine whether each of the following quantities is a matrix, vector, scalar, or nonsense.

a) Mv

$[\;]_{m \times \cancel{n}} [\;]_{\cancel{n} \times 1} = [\;]_{m \times 1}$ vector in $\mathbb{R}^m$

b) vM

$[\;]_{n \times 1} \neq [\;]_{m \times n}$ nonsense

c) v^2

nonsense $\quad n \times 1 \neq n \times 1 \quad v^2 \neq v \cdot v$

d) $M^T M$

$n \times n$ matrix

e) MM^T

$m \times m$ matrix

f) $v^T M v$

$v^T \; M \; v$

$[\;]_{1 \times n} \neq [\;]_{m \times n} [\;]_{n \times 1}$ nonsense

g) $(sMv)\cdot(sMv)$

in $\mathbb{R}^m$ in $\mathbb{R}^m$

vec, vec (of same size) $\rightarrow$ dot prod $\rightarrow$ scalar

h) $(sv^TM^T)^T$

vector in $\mathbb{R}^m$

M is $m \times n$

V is in $\mathbb{R}^n$ (so $n \times 1$)

$$sv^TM^T = s\underbrace{\begin{bmatrix} \quad v^T \quad \end{bmatrix}}_{1\times n}\begin{bmatrix} M^T \end{bmatrix} \quad n\times m$$

$$(s\cdot v^TM^T)^T = (M^T)^T\cdot(v^T)^T\cdot s^T = \underbrace{Mv}_{\text{vector in } \mathbb{R}^m}s$$

i) v^TM^TMv

scalar

$$\underbrace{(Mv)^T}_{1\times m}\underbrace{Mv}_{m\times 1} = (Mv)\cdot(Mv) \quad 1\times m$$

$$\begin{bmatrix} \quad \end{bmatrix} \quad 1\times m$$

j) $vv^T + M^TM$

$n\times n$ matrix

not the same as v^Tv — scalar $= v\cdot v$

$$\underbrace{\begin{bmatrix} v \end{bmatrix}}_{n\times 1}\underbrace{\begin{bmatrix} v^T \end{bmatrix}}_{1\times n} \quad n\times n \text{ matrix}$$

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$$

Problem 2. Orthogonality

Two vectors are **orthogonal** if their dot product is 0, i.e. for $\vec{a}, \vec{b} \in \mathbb{R}^n$:

$$\vec{a}\cdot\vec{b} = \quad \vec{a}^T\vec{b} = 0 \implies \vec{a}, \vec{b} \text{ are orthogonal}$$

Orthogonality is a generalization of perpendicularity to multiple dimensions. (Two orthogonal vectors in 2D meet at a right angle.)

a) Is it possible for a vector to be orthogonal to itself?

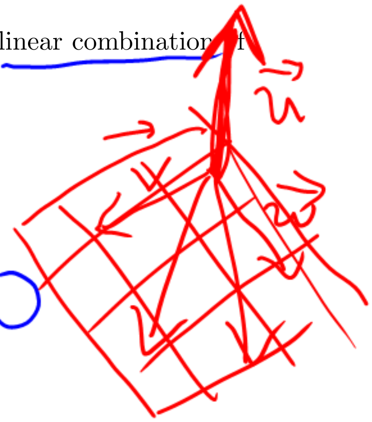
$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \qquad \vec{v}^T\vec{v} = \vec{v}\cdot\vec{v} = 0$$

$$\sum_{i=1}^{n} v_i^2 = v_1^2 + v_2^2 + \dots + v_n^2 = 0$$

$$\vec{0}$$

b) Show that if $\vec{u}$ is orthogonal to both $\vec{v}$ and $\vec{w}$, then $\vec{u}$ is also orthogonal to any linear combination of $\vec{v}$ and $\vec{w}$, $\alpha\vec{v} + \beta\vec{w}$.

Given $u^T v = 0$, $u^T w = 0$.



Show $u^T(\alpha v + \beta w) = 0$

$u^T \alpha v + u^T \beta w$

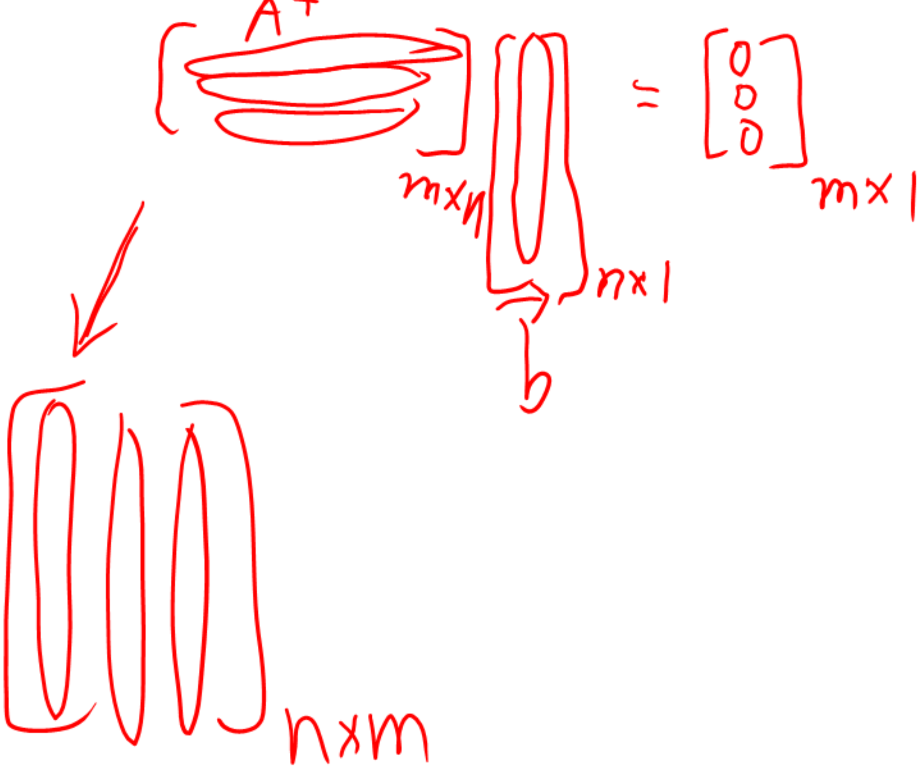
$\alpha u^T v + \beta u^T w$

$\alpha \cdot 0 + \beta \cdot 0 = 0$

$x \cdot 3 \cdot y$

$3 \cdot x \cdot y$

c) Show that if $A^T\vec{b} = 0$, then $\vec{b}$ is orthogonal to the **column space** of A, which is the space of all linear combinations of the columns of A.



(each row of A^T) $\cdot\, b = 0$

(each col of A) $\cdot\, b = 0$

$\vec{b}$ orthogonal to each col. of A

$\Rightarrow \vec{b}$ orthogonal to all linear combos of cols of A

Problem 3. Farmfluencer

Billy the avocado farmer heard about the success of 72 year-old Gerald Stratford's viral gardening videos on Twitter and Instagram. After witnessing Gerald turn into the so-called King of Big Veg overnight, Billy is feeling inspired to up his social media game (he's also feeling a little bit jealous).

Billy is new to Instagram and is trying to understand how people gain followers. In particular, he wants to be able to predict the number of followers, y, based on these features:

- number of people they follow, $x^{(1)}$
- number of years since first post, $x^{(2)}$
- average number of posts per day, $x^{(3)}$

a) Suppose Billy has access to a large data set of Instagram accounts, and he uses multiple regression on this data to fit a linear prediction rule of the form

(+1)

$$y = H(\vec{x}) = w_0 + w_1 x^{(1)} + w_2 x^{(2)} + w_3 x^{(3)}.$$

What does w_2 represent in terms of Instagram followers?

followers for each additional year

"followers per year"

b) What if instead of the number of years since the first post, $x^{(2)}$, Billy instead uses the number of days since the first post, $x^{(4)}$. Now he uses multiple regression to fit a prediction rule of the form

$$H'(\vec{x}) = w_0' + w_1' x^{(1)} + w_3' x^{(3)} + w_4' x^{(4)}.$$

How do the parameters of this prediction rule (w_0', w_1', w_3', w_4') compare to the parameters of original prediction rule (w_0, w_1, w_2, w_3)?

$$\underbrace{365 * w_4'}_{\text{followers/day}} = \underbrace{w_2}_{\text{followers/yr}}$$

How do you identify which feature is most important when features are measured on different scales?

standard units