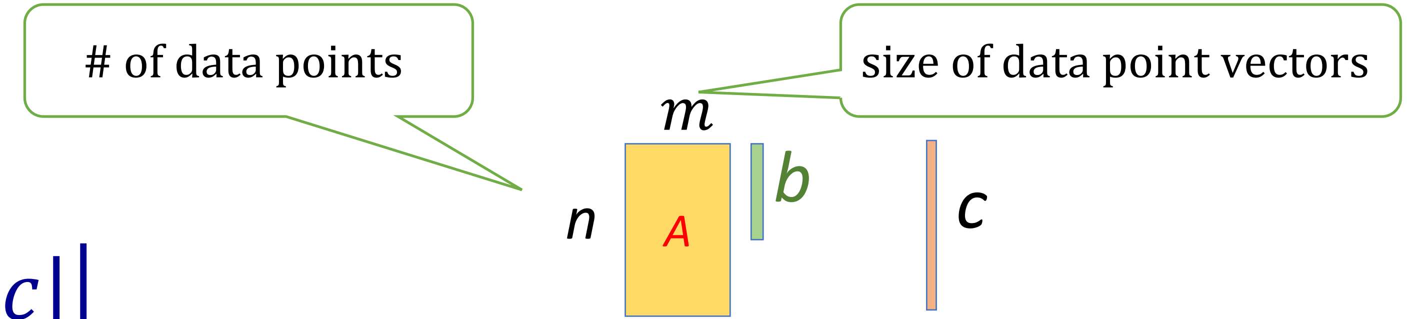
Networked Life

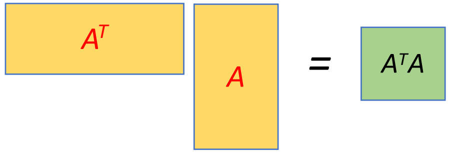
Q4 Tutorial

- Matrix calculus
- Applications

Optimization



- We want to solve $\min_b ||Ab - c||_2$
- This is a convex problem (quadratic in b)
- olve for b the **consistent** set of equations $(A^T A)b = A^T c$
- If A has independent columns, then $A^T A$ is invertible and there is a unique solution $b^* = (A^T A)^{-1} A^T c$

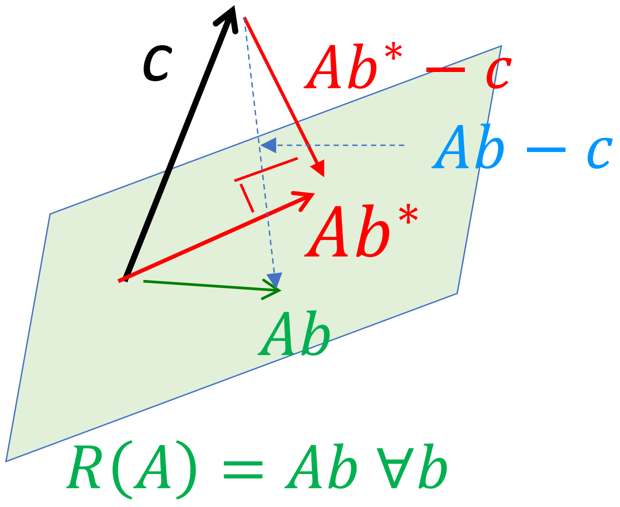


We provide two proofs in the next slides
a) by linear algebra
b) by matrix calculus

Proof A

(*), (**) based on https://ocw.mit.edu/courses/mathematics/18-06sc-linear-algebra-fall-2011/least-squares-determinants-and-eigenvalues/orthogonal-vectors-and-subspaces/MIT18_06SCF11_Ses2.1sum.pdf

- $\min_b ||Ab - c||_2$ is equivalent to $(A^T A)b = A^T c$

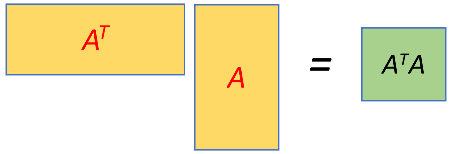


Minimize the distance of c and Ab over all possible bs: project c on the linear subspace range of $A = R(A) \Leftrightarrow$ find b^* s.t. $Ab^* - c \perp R(A)$

$\Leftrightarrow Ab^* - c \in N(A^T)\ \ (*)$

$\Leftrightarrow A^T(Ab^* - c) = 0$

- If A has independent columns, then $A^T A$ is invertible and there is a unique solution $b^* = (A^T A)^{-1} A^T c \quad (**)$



Proof B: matrix calculus

The key matrix calculus properties

scalars = italics

$$a = \mathrm{y}^T \mathrm{x} = \mathrm{x}^T \mathrm{y} \Leftrightarrow \frac{da}{d\mathrm{x}} = (\frac{da}{dx_1}, \cdots, \frac{da}{dx_n}) = \mathrm{y}^T \qquad (*)$$

$$\mathrm{y} = \mathrm{Bx} \Leftrightarrow \frac{d\mathrm{y}}{d\mathrm{x}} = \frac{d}{d\mathrm{x}}(y_1(\mathrm{x}), \ldots, y_n(\mathrm{x}))^T = \mathrm{B} \qquad (**)$$

$$a = \mathrm{y}(\mathrm{x})^T \mathrm{z}(\mathrm{x}) \Leftrightarrow \frac{da}{d\mathrm{x}} = \mathrm{y}(\mathrm{x})^T \frac{d\mathrm{z}(\mathrm{x})}{d\mathrm{x}} + \mathrm{z}(\mathrm{x})^T \frac{d\mathrm{y}(\mathrm{x})}{d\mathrm{x}} \qquad (***)$$

$$a = \mathrm{x}^T \mathrm{Bx} \Leftrightarrow \frac{da}{d\mathrm{x}} \overset{(***)}{=} \mathrm{x}^T \frac{d(\mathrm{Bx})}{d\mathrm{x}} + (\mathrm{Bx})^T \frac{d\mathrm{x}}{d\mathrm{x}} \overset{(**)}{=} \mathrm{x}^T (\mathrm{B} + \mathrm{B}^T) \qquad (1)$$

For matrix calculus please see section 5 in
http://www.atmos.washington.edu/~dennis/MatrixCalculus.pdf

Minimizing SE (calculation to practice matrix calculus)

$$\min_b ||Ab - y||^2 = \min_b (Ab - y)^T (Ab - y)$$

Hence need to find b s.t. $\frac{d}{db} ||Ab - y||^2 = 0$

$$||Ab - y||^2 = (Ab - y)^T (Ab - y) = b^T A^T Ab - y^T Ab - b^T A^T y + y^T y$$

$$\frac{d}{db} ||Ab - y||^2 = b^T \frac{d}{db} (A^T Ab) + (A^T Ab)^T \frac{d}{db} b - y^T A \frac{d}{db} - (A^T y)^T \frac{d}{db} b + 0$$

$$= b^T A^T A + b^T A^T A - y^T A - y^T A$$

$$= 2(b^T A^T - y^T)A = 2(Ab - y)^T A$$

Hence $\frac{d}{db} ||Ab - y||^2 = 0 \Leftrightarrow (Ab - y)^T A = 0$

$$\Leftrightarrow A^T (Ab - y) = 0 \Leftrightarrow A^T Ab = A^T y$$

Minimizing SE (simpler derivation)

$$\min_b ||Ab - y||^2 = \min_b (Ab - y)^T (Ab - y)$$

Hence need to find b s.t. $\frac{d}{db}(Ab - y)^T (Ab - y) = 0$

$$\frac{d}{db}(Ab - y)^T (Ab - y) = (Ab - y)^T \frac{d}{db}(Ab - y) + (Ab - y)^T \frac{d}{db}(Ab - y)$$

$$= 2(Ab - y)^T A$$

$$\frac{d}{db}||Ab - y||^2 = 0 \Leftrightarrow 2(Ab - y)^T A = 0$$

$$\Leftrightarrow A^T (Ab - y) = 0$$

$$\Leftrightarrow A^T A b = A^T y$$

Latent factor example (next 2 slides)

1. Small data set case: over fitting
   - Number of variables (6) > number of ratings (5)
   - We can match the training set exactly!
2. Larger data set case
   - Number of variables (8) = number of ratings (8)
   - Cannot match exactly the ratings since it would require some of the variables to be < 0. Hence, we do not over fit (we see better results for our forecasted ratings for the test set)