Week 11 Cribsheet

NOTES: 23

LEAST SQUARES

super useful in Stats, ML, and their applications

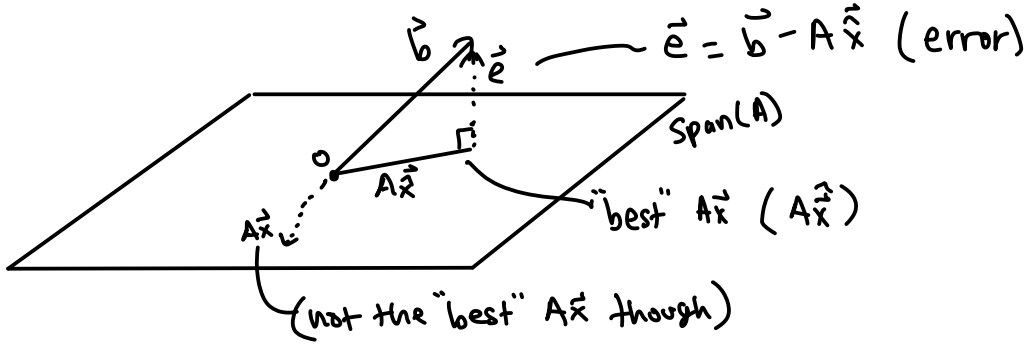
when to use: 1) overdetermined systems

- more equations than unknowns
- real world: more samples than features

2) to deal with noise and errors

Given system $A\vec{x} = \vec{b}$ where A and $\vec{b}$ are known,

approximate $\vec{b}$ with $\hat{b} = A\hat{x}$:



Span(A): by defn, all the linear combinations of A's cols

another interpretation: if we vary $\vec{x}$, all values of $A\vec{x}$

the best $A\hat{x}$ must be the projection of $\vec{b}$ onto span(A):

Formula: $A^TA\hat{x} = A^T\vec{b}$ "normal equations"

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

Requirement: A has linearly independent columns