Quizoops

Is there other notation for least-squares problems?

$\cancel{Ax = b}$ $\qquad\qquad\qquad$ $Ax \cong b$

$\uparrow$
tall + skinny

$\begin{pmatrix}2\\1\end{pmatrix} \cdot x \underset{\cong}{\neq} \begin{pmatrix}4\\5\end{pmatrix}$ $\qquad$ $\left\| \begin{pmatrix}2\\1\end{pmatrix} \cdot x - \begin{pmatrix}4\\5\end{pmatrix} \right\|_2^2$

as small as possible

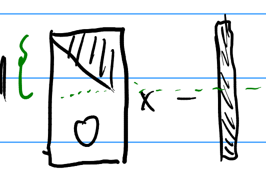
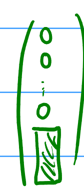
And how does QR help with least-squares problems?

$A = QR$ $\qquad$ A $\quad \overset{\#\text{rows}}{m} \times n$ $\qquad m > n$

$$\|Ax - b\|_2^2 = \|QRx - b\|_2^2$$

$$= \|Q^T(QRx - b)\|_2^2$$

$$= \|Rx - Q^Tb\|_2^2$$

$= \|$  $\|_2^2$ $\quad$ for solution $\hat{x}$ $\leftarrow$ $\quad = \|$  $\|_2^2$

$= \|(Q^Tb)_{bottom}\|_2^2$

↪ R_{top} – first $n \times n$ snippet of R

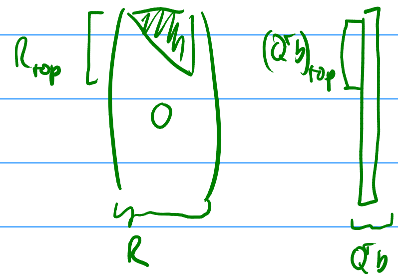
↪ $(Q^T b)_{top}$ – first n entries of $Q^T b$

↪ solve $R_{top}\hat{x} = (Q^Tb)_{top}$

So how do I solve a least-squares problem with QR?

(1) Compute a QR factorization A = QR

(2) Introduce notation



(3) Solve $R_{top} x = (Q^T b)_{top}$ (backsubst)

(4) The norm of the residual can be found as

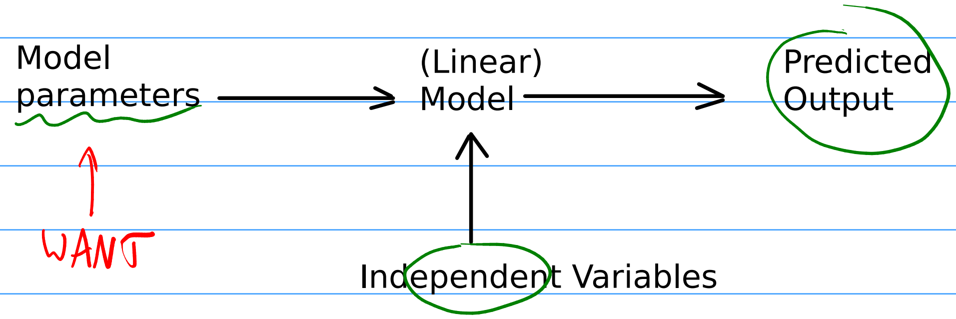
$$\|Ax-b\|_2^2 = \|(Q^T b)_{bottom}\|_2^2$$

What about the "normal equations"?

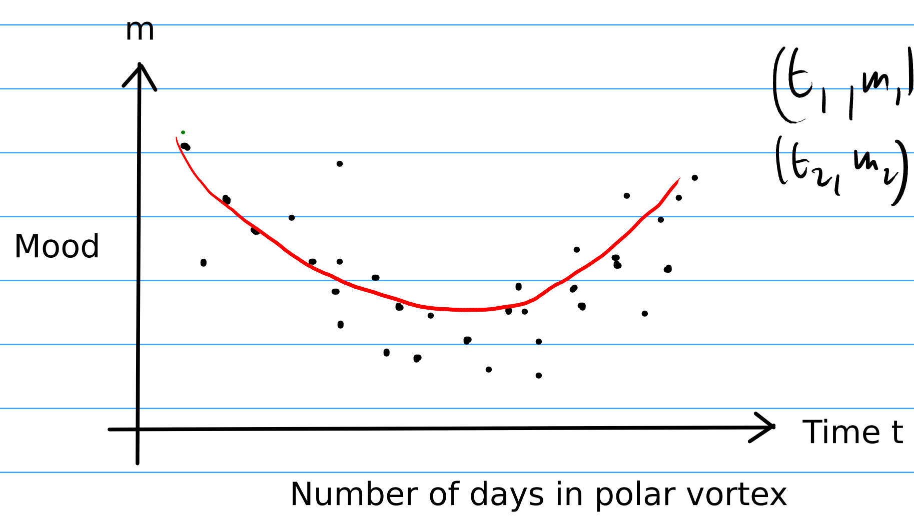
$A^T A x = A^T b$ ← DO NOT USE

$$\kappa(A^T A) \approx \kappa(A)^2$$

How can I use least-squares problems for fitting a linear model to data?



Can you give an example?



(t_1, m_1)
(t_2, m_2)

Model: $\hat{m}(t) = \alpha t^2 + \beta t + \gamma$

How do we find the parameters then?

Have: 30 data points

Want: 3 parameters α, β, γ

Write down equations:

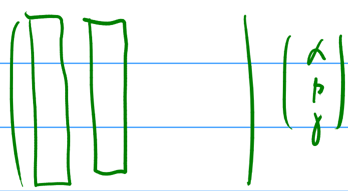
$$\hat{m}(t_1) = \alpha t_1^2 + \beta t_1 + \gamma \approx m_1$$
$$\hat{m}(t_2) = \alpha t_2^2 + \beta t_2 + \gamma = m_2$$
$$\hat{m}(t_3) = \alpha t_3^2 + \beta t_3 + \gamma = m_3$$
$$\vdots$$
$$\hat{m}(t_{30}) = \alpha t_{30}^2 + \beta t_{30} + \gamma = m_{30}$$

$$\rightsquigarrow \begin{pmatrix} t_1^2 & t_1 & 1 \\ t_2^2 & t_2 & 1 \\ t_3^2 & t_3 & 1 \\ & \vdots & \\ t_{30}^2 & t_{30} & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} \simeq \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ \vdots \\ m_{30} \end{pmatrix}$$

If I used a model like

$$\hat{m}(t) = t^{\alpha} + \beta t + \gamma \rightarrow$$

non-linear model.
no straightforward way
to rewrite into matrix-vector
form.



7 Eigenvalues

Remind me: What are eigenvalues and eigenvectors again?

Are there other ways of saying the same thing?

Are there matrices for which the eigenvalues/vectors are easy to find?

What happens to eigenvalues if we change the matrix?

Can we change the eigenvectors? (but leave the eigenvalues the same)

Does every $n \times n$ matrix have n linearly independent eigenvectors?

Can you give an example?

How can we estimate an eigenvalue, given an eigenvector?

So how do we actually go about finding eigenvalues/vectors?

So how do we actually go about finding eigenvalues/vectors? (continued)

Is power iteration bulletproof?

What variants of power iteration are there?

For each variant, you pick an initial, non-zero vector x_0 arbitrarily.

"Plain" power iteration:

$$x_{k+1} = A x_k$$ (overflows quickly if $|\lambda_1| > 1$)

Normalized power iteration:

$$x_{k+1} = \frac{A x_k}{\|A x_k\|}$$ (typically converges to eigenvector for λ_1)

Inverse iteration:

$$x_{k+1} = \frac{A^{-1} x_k}{\|A^{-1} x_k\|}$$ (typically converges to eigenvector for λ_n)

Inverse iteration with shift:

$$x_{k+1} = \frac{(A - \sigma I)^{-1} x_k}{\|(A - \sigma I)^{-1} x_k\|}$$ (typically converges to eigenvector with eigenvalue closest to σ)

greek letter sigma (not zero)

Rayleigh quotient iteration:

Same as Inverse Iteration with Shift, but set σ equal to the Rayleigh quotient computed with x_k.

(Typically converges to one of the eigenvectors, often the one that x_0 is closest to. It does converge very quickly.)

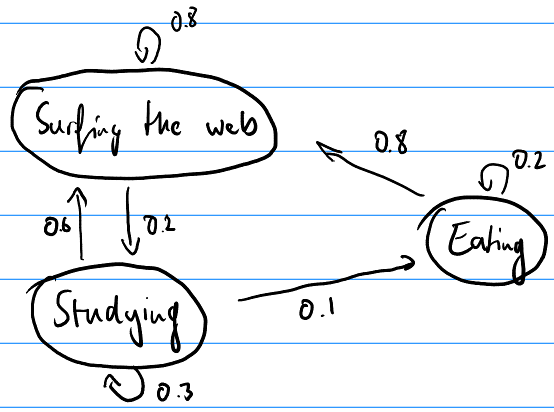
What if we would like to find all eigenvectors at the same time?

How does orthogonal iteration work?

How does that help?

Can we find eigen*vectors* from the Schur form as well?

In what way can eigenvalue computation be applied to Markov chains?



Important Assumption: Only the most recent state matters to determine the probabilities for the next state.

(This is called the "Markov property.")

Write the transition probabilities into a matrix as before.

from state

$$A = \begin{array}{c} \\ \\ \\ \end{array}\begin{array}{ccc} \text{surf} & \text{study} & \text{eat} \\ \end{array}$$

$$A = \begin{pmatrix} .8 & .6 & .8 \\ .2 & .3 & 0 \\ 0 & .1 & .2 \end{pmatrix}\begin{array}{l} \text{surf} \\ \text{study} \\ \text{eat} \end{array} \quad \text{to state}$$

The columns add up to 1 because in each state the probabilities of the next states must add up to 1.

State transitions are modeled by matrix vector product, where the ith vector entry corresponds to the probability of being in the ith state.

$$A\underbrace{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}}_{\text{surf}} = \begin{pmatrix} .8 \\ .2 \\ 0 \end{pmatrix} \qquad A\begin{pmatrix} .5 \\ .5 \\ 0 \end{pmatrix} = \begin{pmatrix} .4 + .3 \\ .1 + .15 \\ .05 \end{pmatrix}$$

"Equilibrium" means that the previous state equals the next state, i.e. the probabilities from one state to the next do not change:

$Ax = \lambda x$ $\rightarrow$ eigenvalue problem

Demo: Finding an equilibrium distribution using the power method