Least Squares Approximation

Rohit Budhiraja, IITK

Applied Linear Algebra for Wireless Communications

Recap and agenda for today's class

- Discussed the following in last lecture
  - Concepts of orthogonal subspaces,and projection today
- Discuss the concept of least squares, and orthonormal matrices today
  - Chapters 4.3, 4.4 of the book

Least Squares Approximations (1)

- Application of least squares is to fit a straight line $b = c + mt$ to k points
- We start with three points: find a line which passed through points
  - $(b_1, t_1) = (6, 0), (b_2, t_2) = (0, 1)$, and $(b_3, t_3) = (0, 2)$
- We can see that no straight line goes through these three points
- We are asking for two numbers m and c that satisfy three equations

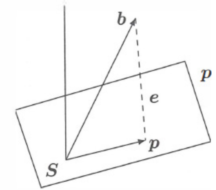
$$\begin{aligned} 6 &= c + m0 \\ 0 &= c + m1 \\ 0 &= c + m2 \end{aligned}$$

- We can equivalently say that 3 by 2 system $A\mathbf{x} = \mathbf{b}$ has no solution:

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \mathbf{x} = \begin{bmatrix} c \\ m \end{bmatrix} \mathbf{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$$

Least Squares Approximations (2)

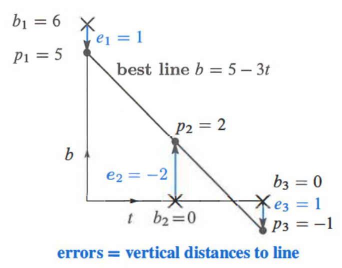
- $A\mathbf{x} = \mathbf{b}$ has no solution $\Rightarrow$ $\mathbf{b}$ does not lie in the column space of A



- Columns $(1, 1, 1)$ and $(0, 1, 2)$ of A span the columns space
- In this plane, we look for "the" point closest to $\mathbf{b}$, which is projection $\mathbf{p} = A\hat{\mathbf{x}}$
  - Projection will have the minimum error $\mathbf{e} = \mathbf{b} - \mathbf{p}$
- We solve this equation $A\hat{\mathbf{x}} = \mathbf{p}$, (which is solvable) instead of $A\mathbf{x} = \mathbf{b}$
  - We know $\hat{x} = (A^T A)^{-1} A^T \mathbf{b}$
  - For this system, we calculated $\hat{\mathbf{x}}$ in last class, is $(5, -3)$
  - $p = A\hat{\mathbf{x}} = (5, 2, -1)$

Least Squares Approximations (3)

- Recall $\hat{\mathbf{x}} = \begin{bmatrix} c \\ m \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$ with the line $b = c + mt$
- So the best line is $b = 5 - 3t$
  - Recall our points are $(b_1, t_1) = (6, 0), (b_2, t_2) = (0, 1)$, and $(b_3, t_3) = (0, 2)$
  - $p = A\hat{\mathbf{x}} = (5, 2, -1)$ and error $e = (1, -2, 1)$



Orthonormal vectors and matrices

- The vectors $q_1, \ldots, q_n$ are orthonormal if

$$\mathbf{q}_i^T \mathbf{q}_j = \begin{cases} 0 \text{ when } i \neq j \text{ (orthogonal vectors)} \\ 1 \text{ when } i = j \text{ (unit vectors : } ||q_1|| = 1) \end{cases}$$

- A matrix with orthonormal columns is assigned the special letter Q
- Matrix Q is easy to work with because $Q^T Q = I$

$$Q^T Q = \begin{bmatrix} -\mathbf{q}_1^T- \\ -\mathbf{q}_2^T- \\ \vdots \\ -\mathbf{q}_n^T- \end{bmatrix} \begin{bmatrix} | & | & \cdots & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ | & | & \cdots & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \mathbf{I}$$

Orthonormal matrices

- Q in general is a rectangular matrix and $Q^TQ = I$, Q^T is the left inverse of Q
- When Q is square, $Q^TQ = I$ means that $Q^T = Q^{-1}$ i.e., Transpose= inverse
- $Q\mathbf{x}$ has same length as $\mathbf{x}$ i.e., $||Q\mathbf{x}|| = ||\mathbf{x}||$
- Proof: $||Q\mathbf{x}||^2 = (Q\mathbf{x})^TQ\mathbf{x} = \mathbf{x}^TQ^TQ\mathbf{x} = \mathbf{x}^T\mathbf{x} = ||\mathbf{x}||^2$
- Orthonormal matrix Q preserves the dot product $(Q\mathbf{x})^TQ\mathbf{y} = \mathbf{x}\mathbf{y}$
- Proof: $(Q\mathbf{x})^TQ\mathbf{y} = \mathbf{x}^TQ^TQ\mathbf{y} = \mathbf{x}\mathbf{y}$

Projection using Orthonormal matrices (1)

- Recall that we have $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$ and $\mathbf{p} = A\hat{\mathbf{x}}$
- If instead of A,we have Q then $\widehat{\mathbf{x}} = Q^T \mathbf{b}$
- There are no matrices to invert. This is the point of an orthonormal basis
- For $\hat{\mathbf{x}}$, we only need to calculate dot products

$$\widehat{\mathbf{x}} = \begin{bmatrix} -\mathbf{q}_1^T- \\ -\mathbf{q}_2^T- \\ \vdots \\ -\mathbf{q}_n^T- \end{bmatrix} \begin{bmatrix} \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1^T\mathbf{b} \\ \mathbf{q}_2^T\mathbf{b} \\ \vdots \\ \mathbf{q}_n^T\mathbf{b} \end{bmatrix} \quad \text{(dot products)}$$

Projection using Orthonormal matrices (3)

- For Projection $\mathbf{p} = Q\hat{\mathbf{x}} = QQ^T\mathbf{b}$

$$\mathbf{p} = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ | & | & \cdots & | \end{bmatrix} \begin{bmatrix} \mathbf{q}_1^T\mathbf{b} \\ \mathbf{q}_2^T\mathbf{b} \\ \vdots \\ \mathbf{q}_n^T\mathbf{b} \end{bmatrix} = \mathbf{q}_1(\mathbf{q}_1^T\mathbf{b}) + \mathbf{q}_2(\mathbf{q}_2^T\mathbf{b}) + \cdots \mathbf{q}_n(\mathbf{q}_n^T\mathbf{b})$$

- When Q is square then $Q^T = Q^{-1}$ and $\mathbf{x} = Q^T\mathbf{b}$ is same as $\mathbf{x} = Q^{-1}\mathbf{b}$
- In this case $\mathbf{p} = \mathbf{b}$ and $P = QQ^T = I$. Projection of $\mathbf{b}$ is $\mathbf{b}$ itself
  - For square Q, column space is the whole space itself
- We may think that projection onto the whole space is not worth mentioning

Projection using Orthonormal matrices (4)

- If $\mathbf{q}_1, \ldots, \mathbf{q}_n$ are orthonormal vectors which form columns of Q then
  - $\mathbf{q}_1, \ldots, \mathbf{q}_n$ are orthonormal basis for $\mathbf{R}^n$ and
  - projection $\mathbf{p} = QQ^T\mathbf{b} = \mathbf{b}$ is the sum of its components along the $\mathbf{q}$'s:

$$\mathbf{b} = \mathbf{q}_1(\mathbf{q}_1^T\mathbf{b}) + \mathbf{q}_2(\mathbf{q}_2^T\mathbf{b}) + \cdots \mathbf{q}_n(\mathbf{q}_n^T\mathbf{b})$$

- Example

$$Q = \frac{1}{3}\begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix} \text{ has first column } \mathbf{q}_1 = \begin{bmatrix} \frac{-1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}$$

- Columns of this orthogonal Q are orthonormal vectors $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$

Projection using Orthonormal matrices (5)

- Separate projections of $\mathbf{b} = (0,0,1)$ onto $\mathbf{q}_1$ $\mathbf{q}_2$ and $\mathbf{q}_3$ are $\mathbf{p}_1$, $\mathbf{p}_2$ and $\mathbf{p}_3$

$$\mathbf{q}_1(\mathbf{q}_1^T\mathbf{b}) = \frac{2}{3}\mathbf{q}_1 \text{ and } \mathbf{q}_2(\mathbf{q}_2^T\mathbf{b}) = \frac{2}{3}\mathbf{q}_2 \text{ and } \mathbf{q}_3(\mathbf{q}_3^T\mathbf{b}) = \frac{-1}{3}\mathbf{q}_3$$

- Sum of first two is the projection of $\mathbf{b}$ onto the plane of $\mathbf{q}_1$ $\mathbf{q}_2$
- Sum of all three is the projection of $\mathbf{b}$ onto the whole space
  - which is $\mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_3 = \mathbf{b}$ itself

$$\frac{2}{3}\mathbf{q}_1 + \frac{2}{3}\mathbf{q}_2 + \frac{-1}{3}\mathbf{q}_3 = \frac{1}{9}\begin{bmatrix} -2+4-2 \\ 4-2-2 \\ 4+4+1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \mathbf{b}$$