Math 415 - Lecture 24

Least squares

Wednesday October 21st 2015

Textbook reading: Chapter 3.3

Suggested practice exercises: Exercises 3, 5, 6, 13, 24, 25

Khan Academy video: Least Squares Approximation, Least Squares Examples, Another Least Squares Example

1 Review

Let W be a subspace of $\mathbb{R}^n$ and $\mathbf{x}$ in $\mathbb{R}^n$ (but maybe not in W). Let $\mathbf{x}_W$ be the orthonormal projection of $\mathbf{x}$ onto W. (vector in W as close as possible to $\mathbf{x}$)

- If $\mathbf{v}_1, \cdots, \mathbf{v}_m$ is an orthogonal basis of W then

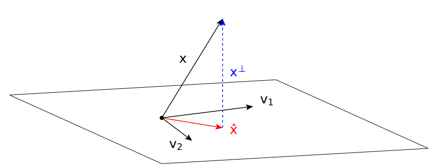
$$\mathbf{x}_W = \underbrace{\left(\frac{\mathbf{x} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right) \mathbf{v_1}}_{\text{proj. of } \mathbf{x} \text{ onto } \mathbf{v}_1} + \cdots + \underbrace{\left(\frac{\mathbf{x} \cdot \mathbf{v}_m}{\mathbf{v}_m \cdot \mathbf{v}_m}\right) \mathbf{v_m}}_{\text{proj. of } \mathbf{x} \text{ onto } \mathbf{v}_m}.$$

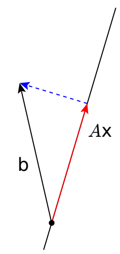
- The decomposition $\mathbf{x} = \underbrace{\mathbf{x}_W}_{\text{in } W} + \underbrace{\mathbf{x}^\perp}_{\text{in } W^\perp}$ is unique.

2 Least squares

2.1 Least squares

Definition. $\hat{\mathbf{x}}$ is a **least squares solution** of the system $A\mathbf{x} = \mathbf{b}$ if $\hat{\mathbf{x}}$ is such that $A\hat{\mathbf{x}} - \mathbf{b}$ is as small as possible.





- If $A\mathbf{x} = \mathbf{b}$ is consistent, then a least squares solution $\hat{\mathbf{x}}$ is just an ordinary solution. (in that case, $A\hat{\mathbf{x}} - \mathbf{b} = 0$)
- Interesting case: $A\mathbf{x} = \mathbf{b}$ is inconsistent. (in other words: the system is overdetermined)

Idea. $A\mathbf{x} = \mathbf{b}$ is consistent $\iff$ $\mathbf{b}$ is in $\mathrm{Col}(A)$

So if $A\mathbf{x} = \mathbf{b}$ is inconsistent we

- replace $\mathbf{b}$ with its projection $\hat{\mathbf{b}}$ onto $\mathrm{Col}(A)$,
- and solve $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$.(consistent by construction!)

Example 1. Find the least squares solution to $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & 0 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}.$$

Solution. Note the columns of A are orthogonal. Otherwise, we could not proceed in the same way. Hence the projection of $\hat{\mathbf{b}}$ of $\mathbf{b}$ onto $\mathrm{Col}(A)$ is

$$\hat{\mathbf{b}} = \frac{\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \frac{\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \frac{3}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

We have already solved $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ in the process: $\hat{\mathbf{x}} = \begin{bmatrix} 1/2 \\ 3/2 \end{bmatrix}$. **Question:** What to do when the columns of A are not orthogonal?

3 The normal equations

Theorem 1. $\hat{\mathbf{x}}$ *is a least squares solution of* $A\mathbf{x} = \mathbf{b}$ $\iff$ $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$

Proof. $\iff$ $A\hat{\mathbf{x}} - \mathbf{b}$ is as small as possible $\iff$ $A\hat{\mathbf{x}} - \mathbf{b}$ is orthogonal to $\mathrm{Col}(A)$ $\overset{FTLA}{\iff}$ $A\hat{\mathbf{x}} - \mathbf{b}$ is in $\mathrm{Nul}(A^T)$ $\iff$ $A^T(A\hat{\mathbf{x}} - \mathbf{b}) = \mathbf{0}$ $\iff$ $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$ $\square$

Example 2 (again). Find the least squares solution to $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & 0 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}.$$

Solution.

$$A^T A = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$A^T \mathbf{b} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

The normal equations $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$ are $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \hat{\mathbf{x}} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$. Solving, we find (again) $\hat{\mathbf{x}} = \begin{bmatrix} 1/2 \\ 3/2 \end{bmatrix}$.

Example 3. Find the least squares solution to $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}.$$

What is the projection of $\mathbf{b}$ onto $\mathrm{Col}(A)$?

Note that the columns of A are not orthogonal.

Solution.

$$A^T A = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

$$A^T \mathbf{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

The normal equations $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$ are $\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \hat{\mathbf{x}} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$. Solving, we find $\hat{\mathbf{x}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. The projection of $\mathbf{b}$ onto $\mathrm{Col}(A)$ is $A\hat{\mathbf{x}} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$.

Just to make sure: why is $A\hat{\mathbf{x}}$ the projection of $\mathbf{b}$ onto $\mathrm{Col}(A)$? Because, for a least square solution $\hat{\mathbf{x}}$, $A\hat{\mathbf{x}} - \mathbf{b}$ is as small as possible.

The projection of $\mathbf{b}$ onto $\mathrm{Col}(A)$ is

$$\hat{\mathbf{b}} = A\hat{\mathbf{x}}, \text{ with } \hat{\mathbf{x}} \text{ such that } A^T A\hat{\mathbf{x}} = A^T\mathbf{b}.$$

If A has full column rank, (so the columns of A are independent,) this is

$$\hat{\mathbf{b}} = A(A^T A)^{-1} A^T \mathbf{b}.$$

(In this case $A^T A$ is invertible.)

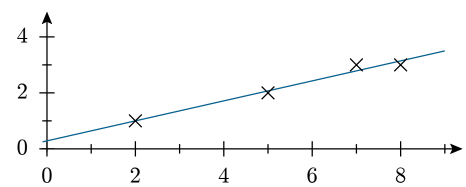
Hence, the projection matrix for projecting onto $\mathrm{Col}(A)$ is

$$P = A(A^T A)^{-1} A^T.$$

4 Applications

4.1 Least square regression lines

Experimental data: (x_i, y_i), for $i = 1, 2, 3, \ldots.$. Wanted: parameters β_1, β_2 such that $y_i \approx \beta_1 + \beta_2 x_i$ for all i



The approximation should be so that

$$\underbrace{\mathrm{SS}_{\mathrm{res}} = \sum_i [y_i - (\beta_1 + \beta_2 x_i)]^2}_{\text{residue sum of squares}} \text{ is as small as possible.}$$

Example 4. Find β_1, β_2 such that the line $y = \beta_1 + \beta_2 x$ best fits the data points $(2, 1), (5, 2), (7, 3), (8, 3)$.

Solution. The equations $y = \beta_1 + \beta_2 x$ in matrix form:

$$\underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix}}_{\text{design matrix } X} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}}_{\text{observation vector } \mathbf{y}}$$

Here, we need to find a least squares solution to

$$\begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}.$$

Calculate X^TX and $X^T\mathbf{y}$ to get the normal equation:

$$X^TX = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

$$X^T\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

Solving $\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \hat{\beta} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$, we find $\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$. Hence the least squares line is $y = \frac{2}{7} + \frac{5}{14}x$.

Example 5. Blood is drawn form volunteers to determine the effects of a new experimental drug designed to lower cholesterol levels. The following data shows the results of varying the dosage from 0 unit to 1 units in step of 0.2 of a unit. Find a line $C = \beta_1 D + \beta_2$ that best fits the data. What drug usage would you recommend if you want to accomplish a Cholesterol level of 215?

Drug Dosage: D	0.0	0.2	0.4	0.6	0.8	1
Cholesterol: C	289	273	254	226	213	189

Solution.

$$\underbrace{\begin{bmatrix} D_1 & 1 \\ D_2 & 1 \\ D_3 & 1 \\ D_4 & 1 \\ D_5 & 1 \\ D_6 & 1 \end{bmatrix}}_{\text{design matrix } D} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \underbrace{\begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \end{bmatrix}}_{\text{observation vector } \mathbf{c}}$$

Here, we need to find a least squares solution to $D\beta = c$ or

$$\begin{bmatrix} 0 & 1 \\ 0.2 & 1 \\ 0.4 & 1 \\ 0.6 & 1 \\ 0.8 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 289 \\ 273 \\ 254 \\ 226 \\ 213 \\ 189 \end{bmatrix}.$$

Calculate D^TD and $D^T\mathbf{c}$ to get the normal equation:

$$D^TD = \begin{bmatrix} 0 & .2 & .4 & .6 & .8 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ .2 & 1 \\ .4 & 1 \\ .6 & 1 \\ .8 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2.2 & 3 \\ 3 & 6 \end{bmatrix}$$

$$D^T\mathbf{c} = \begin{bmatrix} 0 & .2 & .4 & .6 & .8 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 289 \\ 273 \\ 254 \\ 226 \\ 213 \\ 189 \end{bmatrix} = \begin{bmatrix} 651.2 \\ 1444 \end{bmatrix}$$

Solving $\begin{bmatrix} 2.2 & 3 \\ 3 & 6 \end{bmatrix} \hat{\beta} = \begin{bmatrix} 651.2 \\ 1444 \end{bmatrix}$, we find $\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} -\frac{65708}{7} \\ \frac{14116}{21} \end{bmatrix}$. Hence the least squares line is $c = -\frac{65708}{7}d + \frac{14116}{21}$.