

### Study guide

- (online notes) Know what the “least-squares problem” is, in both of its forms: finding the nearest linear combination to a given  $\vec{b}$ , and finding the best approximate solution to an inconsistent system  $A\vec{x} = \vec{b}$ .
- (online notes) Know the “normal equation,” to solve the least-squares problem.
- (online notes) Understand the proof that solutions to the normal equation are necessarily solutions to the least-squares problem.
- (online notes) Understand how to encode linear regression as a least-squares problem, and solve it using the normal equation.

No textbook problems this week.

### Supplemental problems:

1. Prove that if  $\vec{u}$  is a vector that it orthogonal to each vector in a list  $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ , then  $\vec{u}$  is also orthogonal to any linear combination of  $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ .
2. Prove that if  $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$  is a list of vectors, each of which is *nonzero*, and any two of which are orthogonal (that is,  $\vec{v}_i \perp \vec{v}_j$  for all  $i, j$  with  $i \neq j$ ), then  $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$  is a *linearly independent* set.
3. A list of  $m \times 1$  column vectors  $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$  is called an *orthonormal set* if any two of the vectors in the list are orthogonal, and every vector in the list has length 1.
  - (a) Given an index  $i \in \{1, 2, \dots, n\}$ , what do these assumptions say about the dot product  $\vec{v}_i \cdot \vec{v}_i$ ? If  $i$  and  $j$  are two different indices, what do the assumptions say about the dot product  $\vec{v}_i \cdot \vec{v}_j$ ?
  - (b) Suppose that  $\vec{b}$  is another  $m \times 1$  column vector. We can use  $\vec{b}$  to define the following linear combination of  $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ .

$$\vec{u} = (\vec{b} \cdot \vec{v}_1)\vec{v}_1 + (\vec{b} \cdot \vec{v}_2)\vec{v}_2 + \dots + (\vec{b} \cdot \vec{v}_n)\vec{v}_n$$

Prove that if  $\vec{w}$  is any *other* linear combination of  $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ , then  $(\vec{u} - \vec{w}) \perp (\vec{b} - \vec{u})$ .

- (c) Use the Pythagorean theorem for vectors (to be proved in class Monday 10/21) to deduce from part (b) that the vector  $\vec{u}$  is closer to  $\vec{b}$  than any other linear combination of  $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ . (This can also be proved using the normal equation, which takes a particularly simple form when the vectors are orthonormal).
4. Find the linear combination of  $\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$  and  $\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$  that is closest to  $\begin{pmatrix} 5 \\ 3 \\ 7 \end{pmatrix}$ .
  5. Define a matrix  $A$  and vector  $\vec{b}$  as follows.

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad \vec{b} = \begin{pmatrix} 8 \\ 4 \\ 12 \\ 0 \end{pmatrix}$$

- (a) Verify that the linear system  $A\vec{x} = \vec{b}$  is inconsistent.

- (b) Find the “least-squares” solution, i.e. the vector  $\vec{x}$  which minimizes  $\|A\vec{x} - \vec{b}\|$ .
6. Consider the following four points in the plane. This problem will demonstrate a couple ways that we could find a “line of best fit” for these four points. Part (a) is the usual method. The purpose of this exercise is to see how choosing a different “objective” (function to be minimized) can produce different “lines of best fit” (essentially because it depends on what “best” means, which may be different in different applications).

$$(x_1, y_1) = (1, 1) \quad (x_2, y_2) = (3, 2) \quad (x_3, y_3) = (1, 6) \quad (x_4, y_4) = (3, 7)$$

- (a) Suppose that we wish to find the coefficients  $c_1, c_2$  that minimize the sum  $\sum_{i=1}^4 (c_1 x_i + c_2 - y_i)^2$ .

Identify vectors  $\vec{v}_1, \vec{v}_2$ , and  $\vec{b}$  such that this is the same as minimizing  $\|c_1 \vec{v}_1 + c_2 \vec{v}_2 - \vec{b}\|$ . Then find the optimal coefficients  $c_1, c_2$ . Sketch the four points and the line  $y = c_1 x + c_2$ .

- (b) Suppose that we now want coefficients  $c_1, c_2$  that minimize  $\sum_{i=1}^4 (c_1 y_i + c_2 - x_i)^2$ . Identify

vectors  $\vec{v}_1, \vec{v}_2$ , and  $\vec{b}$  such that this is the same as minimizing  $\|c_1 \vec{v}_1 + c_2 \vec{v}_2 - \vec{b}\|$ . Then find the optimal coefficients  $c_1, c_2$  and sketch the four points and the line  $x = c_1 y + c_2$ .

- (c) A third way to specify a line is using an equation of the form  $c_1 x + c_2 y = 1$ . Suppose

that now we wish to find  $c_1, c_2$  minimizing  $\sum_{i=1}^4 (c_1 x_i + c_2 y_i - 1)^2$ . Identify vectors  $\vec{v}_1, \vec{v}_2$ ,

and  $\vec{b}$  such that this is the same as minimizing  $\|c_1 \vec{v}_1 + c_2 \vec{v}_2 - \vec{b}\|$ . Then find the optimal coefficients  $c_1, c_2$  and sketch the four points with the line  $c_1 x + c_2 y = 1$ .

*Comment:* There is another objective function that is often used in practice to find lines (or more generally, linear spaces) of best fit: one can write the line as  $c_1 x + c_2 y = c_3$ , and attempt to minimize the sum  $\sum_{i=1}^4 (c_1 x_i + c_2 y_i - c_3)^2$ . As stated, this problem has a silly solution: choose  $c_1 = c_2 = c_3 = 0$ . One can fix this by requiring that  $c_1^2 + c_2^2 + c_3^2 = 1$ . The solution to this problem is related to a topic called *singular value decomposition*, an extremely versatile idea in machine learning and data science. One benefit of this objective function is that it does not arbitrarily prefer horizontal error to vertical error, but rather treats the  $x$  and  $y$  axis on equal terms.