**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 how to encode linear regression as a least-squares problem, and solve it using the normal equation.

1. Find the linear combination of $\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$ that is closest to $\begin{pmatrix} 5 \\ 3 \\ 7 \end{pmatrix}$.

2. Define a matrix $A$ and vector $\vec{b}$ as follows.

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

   (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 $\|A\vec{x} - \vec{b}\|$.

   **Note** This is a bit different from the original way we formulated the least-squares problem. To translate, observe that $A\vec{x}$ is a linear combination of the *columns* of $A$, or check the online notes.

♣ 3. Suppose that $\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\}$ is a linearly independent set of vectors in $\mathbb{R}^n$, and let $\vec{v}_{n+1}$ be some other vector in $\mathbb{R}^n$. Prove that $\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n, \vec{v}_{n+1}\}$ is linearly dependent *if and only if* $\vec{v}_{n+1}$ is a linear combination of the vectors $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n$.

   **Note** Be very careful in your proof that you never assume that a number is nonzero without justifying why it must be so.

♣ 4. A square matrix $A$ is called *orthogonal* if it is invertible and $A^{-1} = A^t$. (Note: read Theorem 15 in §1.6 of the book. We did not explicitly mention all of these facts in class, but a couple of them are useful here.)

   (a) Prove that if $A$ is orthogonal, then $\det A = \pm 1$.

   (b) Prove that if $A$ is an orthogonal $n \times n$ matrix and $\vec{v}$ is any vector in $\mathbb{R}^n$, then $\|A\vec{v}\| = \|\vec{v}\|$.

   (c) Prove that if $A$ is an orthogonal $n \times n$ matrix and $\vec{u}, \vec{v}$ are any two nonzero vectors in $\mathbb{R}^n$, then the angle between $A\vec{u}$ and $A\vec{v}$ is the same as the angle between $\vec{u}$ and $\vec{v}$.

   (d) Prove that the product of two orthogonal matrices of the same size is orthogonal, and that the inverse of an orthogonal matrix is orthogonal.

> **Note** Orthogonal matrices arise in physics and engineering, because they represent rigid motions (rotations, etc.), which is demonstrated by the properties discussed above.

♣ 5. Prove that if $\vec{u}$ is a vector that is orthogonal to each vector in a list $\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\}$, then $\vec{u}$ is also orthogonal to any linear combination of $\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\}$.

♣ 6. Prove that if $\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\}$ 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 $\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\}$ is a *linearly independent* set.

♣ 7. A list of $m \times 1$ column vectors $\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\}$ 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 $i \in \{1, 2 \cdots, n\}$, 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 $\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\}$.

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

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

(c) Use the Pythagorean theorem for vectors to deduce from part (b) that the vector $\vec{u}$ is closer to $\vec{b}$ than any other linear combination of $\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\}$. (This can also be proved using the normal equation, which takes a particularly simple form when the vectors are orthonormal).