Exercises:

1. For any points $\vec{b} \in \mathbb{R}^n$, define $\vec{b}_{proj} = \sum_{j=1}^{r}(\vec{b} \cdot \hat{u}_j)\hat{u}_j$. Show for all $\vec{b}_{proj}$ belongs to the span of the columns of A, and that for all y in the span of the columns of A, $\|\vec{b}_{proj} - \vec{b}\|^2 \leq \|\vec{y} - \vec{b}\|^2$.

   - Show that $\vec{b}_{proj} \in Span\{Cols(A)\}$.
     Since $\vec{b}_{proj} = \sum_{j=1}^{r}(\vec{b} \cdot \hat{u})\hat{u}$, and each $\hat{u}_j$ is a linear combination of the rows of A, then $\vec{b}_{proj}$ is a linear combination of the columns of A.
   - $\|\vec{b}_{proj} - \vec{b}\|^2 \leq \|\vec{y} - \vec{b}\|^2$.
     Suppose $\vec{y} \in Span\{A\}$. Then $\vec{y} - \vec{b} = \vec{y} - \vec{b}_{proj} + \vec{b}_{proj} - \vec{b}$. Since $\vec{b}_{proj} \in Span\{A\}$, and $\vec{b} \notin Span\{A\}$, then $\vec{y} - \vec{b}_{proj}$ is orthogonal to $\vec{b}_{proj} - \vec{b}$. Therefore by the Pythagorean theorem, $\|\vec{y} - \vec{b}\|^2 = \|\vec{y} - \vec{b}_{proj}\|^2 + \|\vec{b}_{proj} - \vec{b}\|^2$. Since $\|\vec{b}_{proj} - \vec{b}\|^2$ is constant then the only function that can be minimized is $\|\vec{y} - \vec{b}_{proj}\|^2$, which defines the square of the distance to $\vec{b}_{proj}$. Therefore it is minimized when $\vec{y} = \vec{b}_{proj}$, thus $\|\vec{b}_{proj} - \vec{b}\|^2 \leq \|\vec{y} - \vec{b}\|^2$.

2. Show that for all $\vec{x} \in \mathbb{R}^n$,
   $$\|A\vec{x}_0 - \vec{b}\|^2 \leq \|A\vec{x} - \vec{b}\|^2.$$
   Let $\vec{y} = A\vec{x}$, for arbitrary $\vec{x} \in \mathbb{R}^n$. Then the above equation may be rewritten as $\|\vec{b}_{proj} - \vec{b}\|^2 \leq \|\vec{y} - \vec{b}\|^2$. Therefore as demonstrated by the problem above the inequality is true.

3. Show that for $s + 1 \leq j \leq n, A\hat{w}_j = 0$. Note that $\{\hat{w}_{s+1}, \ldots, \hat{w}_n\}$ is an orthonormal basis of the null space $Null(A)$ of A. Next show that if $\vec{x}_0$ satisfies $A\vec{x}_0 = \vec{b}_{proj}$, then so does
   $$\vec{x}_1 := \vec{x}_0 - \sum_{j=s+1}^{n} (\mathbf{x}_0 \cdot \vec{w}_j)\, \vec{w}_j = \sum_{j=1}^{s} (\vec{x}_0 \cdot \vec{w}_j)\, \vec{w}_j.$$
   Moreover, show that $\|\vec{x}_1\| \leq \|\vec{x}_0\|$, and there is equality if and only if $\vec{x}_0$ is in the span of the rows of A.

   - Show that for $s + 1 \leq j \leq n, A\hat{w}_j = 0$.
     Since $Span\{A\} = Q$, and Q is given by $Q = [\hat{w}_1, \cdots, \hat{w}_s]$, then $Q^{\perp} = [\hat{w}_{s+1}, \cdots, \hat{w}_n]$. Therefore by the definition of orthogonality $A\hat{w}_j = [\hat{w}_1 \cdot \hat{w}_j, \cdots, \hat{w}_s \cdot \hat{w}_j] = [0, \cdots, 0]$.
   - Show that if $\vec{x}_0$ satisfies $A\vec{x}_0 = \vec{b}_{\text{proj}}$, then so does
     $$\vec{x}_1 := \vec{x}_0 - \sum_{j=s+1}^{n} (\vec{x}_0 \cdot \vec{w}_j)\, \vec{w}_j = \sum_{j=1}^{s} (\vec{x}_0 \cdot \vec{w}_j)\, \vec{w}_j.$$

$$\begin{aligned} A\vec{x}_1 &= A(\vec{x}_0 - \sum_{j=s+1}^{n} (\vec{x}_0 \cdot \vec{w}_j)\, \vec{w}_j) \\ &= A\vec{x}_0 - A \sum_{j=s+1}^{n} (\vec{x}_0 \cdot \vec{w}_j)\, \vec{w}_j \\ &= A\vec{x}_0 - \sum_{j=s+1}^{n} (\vec{x}_0 \cdot \vec{w}_j)\, A\vec{w}_j \\ &= A\vec{x}_0 - 0 = \vec{b}_{proj}. \end{aligned}$$

- Show that $\|\vec{x}_1\| \leq \|\vec{x}_0\|$, and there is equality if and only if $\vec{x}_0$ is in the span of the rows of A
  Since $\vec{x}_0 - \sum_{j=s+1}^{n} (\mathbf{x}_0 \cdot \vec{w}_j)\, \vec{w}_j = \sum_{j=1}^{s} (\vec{x}_0 \cdot \vec{w}_j)\, \vec{w}_j$, then x_0 can be expressed in terms of parallel and orthogonal components with respect to A: $\vec{x}_0 = \sum_{j=s+1}^{n} (\mathbf{x}_0 \cdot \vec{w}_j)\, \vec{w}_j + \sum_{j=1}^{s} (\vec{x}_0 \cdot \vec{w}_j)\, \vec{w}_j$. Therefore by the Pythagorean theorem , $\|\vec{x}_0\|^2 = \| \sum_{j=s+1}^{n} (\mathbf{x}_0 \cdot \vec{w}_j)\, \vec{w}_j\|^2 + \| \sum_{j=1}^{s} (\vec{x}_0 \cdot \vec{w}_j)\, \vec{w}_j\|^2 = \| \sum_{j=s+1}^{n} (\vec{x}_0 \cdot \vec{w}_j)\, \vec{w}_j\|^2 + \|\vec{x}_1\|^2 \geq \|\vec{x}_1\|^2$. Since the norm of a vector has a range of $\mathbb{R}_{\geq 0}$, and x^2 is a monotonically increasing function on that set, then $\|\vec{x}_1\| \leq \|\vec{x}_0\|$. In addition, if $\vec{x}_0$ is in the span of the rows of A, then $\sum_{j=s+1}^{n} (\vec{x}_0 \cdot \vec{w}_j)\, \vec{w}_j = 0$, and thus $\|\vec{x}_0\| = \|\vec{x}_1\|$. If we assume $\|\vec{x}_1\| = \|\vec{x}_0\|$, then by definition of $\vec{x}_1$, $\|\vec{x}_0\| = \| \sum_{j=1}^{s} (\vec{x}_0 \cdot \vec{w}_j)\, \vec{w}_j\| = \|QQ^T\vec{x}_0\|$. Therefore since the magnitude of $\vec{x}_0$ is equivalent to it's magnitude projected into A, then it is in the span of the rows of A.

4. (a) $\vec{b} = (1, 2, 3, 4)$, $\|\vec{b} - \vec{b}_{proj}\| = 1$, unique solution $\vec{x}$ to $A\vec{x} = \vec{b}_{proj}$, $\vec{x} = (2, \frac{-5}{2}, 0, \frac{3}{2})$.
   (b) $\vec{b} = (1, 1, 1, -1)$, $\|\vec{b} - \vec{b}_{proj}\| = 2$, unique solution $\vec{x}$ to $A\vec{x} = \vec{b}_{proj}$, $\vec{x} = (0, 0, 0, 0)$.
   (c) $\vec{b} = (1, 1, -1, 1)$, $\|\vec{b} - \vec{b}_{proj}\| = 0$, unique solution $\vec{x}$ to $A\vec{x} = \vec{b}_{proj}$, $\vec{x} = (1, 0, 0, 0)$.

5. - Show that as long as all of the $\vec{x}_j$ are not the same, then $rank(A) = 2$ and the 2×2 matrix A^TA is invertible. By definition of invertible, we must show that $det(A^TA) \neq 0$. Since $rank(A) = 2$, then when we perform gram schmidt we will get $\{\hat{u}_1, \hat{u}_2\} = Span\{A\}$. Therefore let $Q = \{\hat{u}_1, \hat{u}_2\}$. Then by QR factorization $A^TA = R^TQ^TQR = R^TR$. By definition of R,

$$R = \begin{bmatrix} Col_1(A) \cdot \hat{u}_1 & \cdots & Col_n(A) \cdot \hat{u}_1 \\ Col_1(A) \cdot \hat{u}_2 & \cdots & Col_n(A) \cdot \hat{u}_2 \end{bmatrix}.$$

Therefore

$$R^TR = \begin{bmatrix} \sum_{j=1}^{n}(Col_j(A) \cdot \hat{u}_1)^2 & \sum_{j=1}^{n}(Col_j(A) \cdot \hat{u}_1)(Col_j(A) \cdot \hat{u}_2) \\ \sum_{j=1}^{n}(Col_j(A) \cdot \hat{u}_1)(Col_j(A) \cdot \hat{u}_2) & \sum_{j=1}^{n}(Col_j(A) \cdot \hat{u}_2)^2 \end{bmatrix}$$

$$= \begin{bmatrix} \|A^T\hat{u}_1\|^2 & (A^T\hat{u}_1) \cdot (A^T\hat{u}_2) \\ (A^T\hat{u}_1) \cdot (A^T\hat{u}_2) & \|A^T\hat{u}_2\|^2 \end{bmatrix}.$$

Therefore $det(A^TA) = (\|A^T\hat{u}_1\|\|A^T\hat{u}_2\|)^2 - ((A^T\hat{u}_1) \cdot (A^T\hat{u}_2))^2 = (\|A^T\hat{u}_1\|\|A^T\hat{u}_2\| +$

$(A^T\hat{u}_1)\cdot(A^T\hat{u}_2))(\|A^T\hat{u}_1\|\|A^T\hat{u}_2\| - (A^T\hat{u}_1)\cdot(A^T\hat{u}_2))$. The only way for this determinate to be 0 is if $A^T\hat{u}_1 = A^T\hat{u}_2$. However this implies that the first row of R is equal to the second. In that case R would have a rank of 1, which would contradict the fact that $rank(A) = 2$. Therefore $det(A) \neq 0$, and thus A^TA has an inverse.

- Show for all $\vec{y} \in \mathbb{R}^m, \exists \vec{x} \in \mathbb{R}^n$ such that for all $\vec{z} \in \mathbb{R}^n, \|A\vec{x} - \vec{y}\| \leq \|A\vec{z} - \vec{y}\|$.
  Since $\vec{y} \in \mathbb{R}^m$, then $\vec{y}$ decomposes into parts that are parallel and orthogonal to $A : \vec{y} = \vec{y}_\perp + \vec{y}_\|$. The portion which is parallel to A, $\vec{y}_\|$, by definition can be written as the product of a vector $\vec{x}_0$ in $\mathbb{R}^n$ and $A : \vec{y}_\| = A\vec{x}_0$. Therefore by taking the decomposition of $\vec{y}$ and the formula found for the parallel portion, we can solve for the orthogonal part: $\vec{y} = \vec{y}_\perp + \vec{y}_\| = \vec{y}_\perp + A\vec{x}_0$, $\vec{y} - A\vec{x}_0 = \vec{y}_\perp$. Therefore $\vec{y} - A\vec{z} = \vec{y}_\perp + \vec{y}_\| - A\vec{z}$. Since by definition $\vec{y}_\perp$ is orthogonal to A, and $A\vec{z}$ lies explicitly within A, then by the pythagorean theorem $\|\vec{y} - A\vec{z}\|^2 = \|\vec{y}_\perp\|^2 + \|\vec{y}_\| - A\vec{z}\|^2$. Therefore to minimize the above equation, we must choose $\vec{x} = \vec{x_0}$ so that $\vec{y}_\| - A\vec{z} = \vec{0}$.
- Show that $A\vec{x} - \vec{y} \perp Col_j(A)$, for $j = 1, \cdots, n$.
  Since it's established that $A\vec{x} - \vec{y}$ is orthogonal to the columns of A, then by definition is is orthogonal to each individual columns $Col_j(A)$
- Finally show that $(A^TA)^{-1}A^T\vec{y}$ is the unique least squares solution of $A\vec{x} = \vec{y}$.
  Suppose there exists $\vec{x}^*$ such that $A\vec{x}^* = \vec{y}$ in addition to $\vec{x} = (A^TA)^{-1}A^T\vec{y}$. We must show that $\vec{x}^* = (A^TA)^{-1}A^T\vec{y}$. Since both $A\vec{x}^* = \vec{y}, A\vec{x} = \vec{y}$, then $A\vec{x}^* = A\vec{x}$. By definition of $\vec{x}$, $A\vec{x}^* = A(A^TA)^{-1}A^T\vec{y}$. Multiplying both sides by A^T yields $A^TA\vec{x}^* = A^TA(A^TA)^{-1}A^T\vec{y}$. By the associativity of matrix multiplication and the definition of inverse we have, $A^TA\vec{x}^* = A^T\vec{y}$. Since A^TA has been shown to have an inverse, multiplying both sides by $(A^TA)^{-1}$ yields $\vec{x}^* = (A^TA)^{-1}A^T\vec{y}$.

6. $y = 1.3369x - 1.13274$