ReLU Networks: Comparing Two Functions

QUESTION 01

Consider

$$f(x) = \mathrm{ReLU}(x+1) - 2\,\mathrm{ReLU}(x) + \mathrm{ReLU}(x-1),$$
$$g(x) = 1 - \mathrm{ReLU}(-x-1) - \mathrm{ReLU}(x-1).$$

1. Find all possible kink locations of f and g.
2. Write f and g as piecewise linear functions.
3. Determine whether $f(x) = g(x)$ for all $x \in \mathbb{R}$.
4. Draw a shallow ReLU network for f; specify every weight and bias.

ReLU Networks: Comparing Two Functions

ANSWER 01

Kinks: f: $-1, 0, 1$; g: $-1, 1$.

$$f(x) = \begin{cases} 0, & x \le -1, \\ x+1, & -1 < x \le 0, \\ 1-x, & 0 < x \le 1, \\ 0, & x > 1, \end{cases} \qquad g(x) = \begin{cases} x+2, & x < -1, \\ 1, & -1 \le x \le 1, \\ 2-x, & x > 1. \end{cases}$$

Not equal: e.g., $f(1/2) = 1/2$ but $g(1/2) = 1$. **Network for f:** three hidden ReLU units with

$$(w_1, w_2, w_3) = (1, 1, 1), \quad (b_1, b_2, b_3) = (1, 0, -1), \quad (c_1, c_2, c_3) = (1, -2, 1).$$

Output bias 0.

ReLU Networks and Minimal Width

QUESTION 02

Let

$$f(x) = \min\{|x+1|,\ |x-1|\}, \qquad x \in \mathbb{R}.$$

Represent f by a shallow ReLU network

$$f(x) = d + \sum_{j=1}^{m} c_j\,\mathrm{ReLU}(w_j x + b_j), \qquad d, c_j, w_j, b_j \in \mathbb{R}.$$

(a) Construct a representation using at most four hidden neurons. Give all weights and biases.

(b) Prove that no representation with only two hidden neurons is possible.

ANSWER 02

The function has four linear pieces:

$$f(x) = \begin{cases} -x - 1, & x \le -1, \\ x + 1, & -1 < x \le 0, \\ 1 - x, & 0 < x \le 1, \\ x - 1, & x > 1. \end{cases}$$

(a) One four-neuron construction is

$$\boxed{f(x) = \mathrm{ReLU}(-x - 1) + \mathrm{ReLU}(x + 1) - 2\,\mathrm{ReLU}(x) + 2\,\mathrm{ReLU}(x - 1).}$$

Here $d = 0$, $w = (-1, 1, 1, 1)$, $b = (-1, 1, 0, -1)$, $c = (1, 1, -2, 2)$. **(b)** There are three genuine kinks at $x = -1, 0, 1$. Each ReLU neuron has at most one kink, so two hidden neurons can produce at most two kink locations. Contradiction.

Least-Squares Uniqueness

QUESTION 03

Let $A \in \mathbb{R}^{n \times n}$ be square and suppose

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 > 0.$$

Question. Under what condition on A is the least-squares solution unique? Justify your answer.

Least-Squares Uniqueness

ANSWER 03

General criterion: uniqueness holds if and only if

$$\boxed{\mathcal{N}(A) = \{0\} \quad \text{(full column rank).}}$$

If $\widehat{x}$ is any minimizer, then so is $\widehat{x}+z$ for every $z \in \mathcal{N}(A)$. Therefore uniqueness requires $\mathcal{N}(A) = \{0\}$. **But note the extra assumption in this problem!** Since A is square, $\mathcal{N}(A) = \{0\}$ would make A invertible, giving $Ax = b$ for some x and hence

$$\min_x \|Ax - b\|_2^2 = 0,$$

contrary to the hypothesis. Conclusion: under the stated hypothesis, the least-squares solution is never unique.

Nonunique Least-Squares Solutions

QUESTION 04

Let

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix}, \qquad b = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Suppose $\widehat{x}$ is one least-squares solution to $Ax \approx b$. Without explicitly solving the least-squares problem:

1. Determine whether $\widehat{x}$ is unique.
2. If it is not unique, describe the set of all least-squares solutions.

Nonunique Least-Squares Solutions

ANSWER 04

The third column of A equals the sum of its first two:

$$a_3 = a_1 + a_2.$$

Thus $\text{rank}(A) = 2 < 3$ and

$$\mathcal{N}(A) = \text{span}\left\{\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}\right\}.$$

Not unique. Every $z \in \mathcal{N}(A)$ satisfies $A(\widehat{x} + z) = A\widehat{x}$, so it gives exactly the same residual.

All minimizers: $\widehat{x} + t \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}, \qquad t \in \mathbb{R}.$

True or False?

QUESTION 05

Determine whether each statement is true or false.

(a) If the columns of Q are orthonormal, then $QQ^T = I$.

(b) If $A^TAx = 0$, then $Ax = 0$.

(c) If $\mathcal{N}(A) = \{0\}$, the least-squares solution is unique for every b.

(d) For subspaces L, K, $\text{span}(L \cap K) = \text{span}(L) \cap \text{span}(K)$.

(e) If $A = QR$ is a reduced QR of a rank-r matrix, then $\mathcal{C}(A) = \mathcal{C}(Q)$.

(f) If P is an orthogonal projection matrix, then $P^2 = P$.

(g) If A is symmetric and $x^TAx = 0$ for every $x \in \mathbb{R}^n$, then $A = 0$.

True or False?

ANSWER 05

Statement	Answer
(a)	**False**
(b)	**True**
(c)	**True**
(d)	**True**
(e)	**True**
(f)	**True**
(g)	**True**

Rank of AA^T

QUESTION 06

Let $A \in \mathbb{R}^{m \times n}$. **Prove that**

$$\boxed{\text{rank}(AA^T) = \text{rank}(A).}$$

Rank of AA^T

ANSWER 06

First show that

$$\mathcal{N}(AA^T) = \mathcal{N}(A^T).$$

Certainly, $A^T x = 0 \Rightarrow AA^T x = 0$. Conversely, if $AA^T x = 0$, then

$$0 = x^T AA^T x = \|A^T x\|_2^2,$$

so $A^T x = 0$. By rank-nullity for AA^T and A^T,

$$\begin{aligned} \operatorname{rank}(AA^T) &= m - \dim \mathcal{N}(AA^T) \\ &= m - \dim \mathcal{N}(A^T) \\ &= \operatorname{rank}(A^T) = \boxed{\operatorname{rank}(A)}. \end{aligned}$$

Null Space from AA^T

QUESTION 07

Let $A \in \mathbb{R}^{m \times n}$, and suppose

$$AA^T = \begin{pmatrix} 2 & 2 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Without knowing A itself:

1. Find $\mathcal{N}(A^T)$ and justify your answer.
2. Can you determine $\mathcal{N}(A)$? Justify your answer.
3. Determine $\text{rank}(A)$.

Null Space from AA^T

ANSWER 07

1. Since

$$x^T AA^T x = \|A^T x\|_2^2,$$

we have

$$\mathcal{N}(A^T) = \mathcal{N}(AA^T).$$

Solving $AA^T x = 0$ gives

$$\boxed{\mathcal{N}(A^T) = \text{span}\left\{\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}\right\}.}$$

2. No. Different matrices can have the same AA^T but different null spaces.
For example,

$$A_1 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

They satisfy $A_1 A_1^T = A_2 A_2^T$, but

$$\mathcal{N}(A_1) = \text{span}\{(1, -1, 0)^T\}, \qquad \mathcal{N}(A_2) = \text{span}\{(0, 0, 1)^T\}.$$

Orthogonal Projections onto Nested Subspaces

QUESTION 08

Let $K \subseteq L$ be two subspaces of $\mathbb{R}^n$. Denote by P_K and P_L the orthogonal projections onto K and L, respectively. **Prove that for every $x \in \mathbb{R}^n$,**

$$\boxed{P_K(P_L x) = P_L(P_K x) = P_K x.}$$

Orthogonal Projections onto Nested Subspaces

ANSWER 08

Since $K \subseteq L$, the vector $P_K x$ belongs to L. The projection P_L fixes all vectors in L, hence

$$\boxed{P_L(P_K x) = P_K x.}$$

Next, by the defining property of orthogonal projection,

$$x - P_L x \in L^\perp \subseteq K^\perp.$$

Therefore $P_K(x - P_L x) = 0$. By linearity,

$$P_K x - P_K(P_L x) = 0,$$

so

$$\boxed{P_K(P_L x) = P_K x.}$$

Combining both identities proves the claim.

Least Squares and Minimum-Norm Solutions

QUESTION 09

Consider

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix}, \qquad b = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}.$$

(a) Without computing $A^T A$, determine whether the least-squares solution is unique. Justify.

(b) Find the orthogonal projection of b onto $\mathcal{C}(A)$.

(c) Find all least-squares solutions.

(d) Find the solution with the smallest Euclidean norm.

Least Squares and Minimum-Norm Solutions

ANSWER 09

(a) $a_3 = a_1 + a_2$, so $\operatorname{rank}(A) = 2$: solutions are not unique. **(b)** $\mathcal{C}(A)^{\perp} = \operatorname{span}\{(1,1,-1)^T\}$. Therefore

$$P_{\mathcal{C}(A)}b = b - \frac{b^T(1,1,-1)^T}{3}(1,1,-1)^T = \boxed{\begin{pmatrix}0\\1\\1\end{pmatrix}}.$$

(c) Since $Ax = P_{\mathcal{C}(A)}b$, all solutions are

$$\boxed{x = (-t,\, 1-t,\, t)^T, \qquad t \in \mathbb{R}.}$$

(d) Minimize $\|x\|_2^2 = 3t^2 - 2t + 1$: $t = 1/3$, giving $\boxed{x_{\min} = (-1/3,\, 2/3,\, 1/3)^T}$.

Dimension of Intersections

QUESTION 10

Let L, K be subspaces of $\mathbb{R}^6$ satisfying

$$\dim L = 4, \qquad \dim K = 5.$$

(a) Prove that $\dim(L \cap K) \geq 3$.

(b) Determine all possible values of $\dim(L \cap K)$.

(c) Construct explicit L, K attaining each possible value.

(d) Suppose additionally $L^{\perp} \subseteq K$. Does this uniquely determine $\dim(L \cap K)$? Prove or give a counterexample.

Dimension of Intersections

ANSWER 10

(a) $\dim(L \cap K) = 4 + 5 - \dim(L + K) \geq 9 - 6 = 3$. **(b)** Also $\dim(L \cap K) \leq 4$; thus the only possibilities are $\boxed{\text{3 and 4}}$. **(c)** Let e_i be the standard basis of $\mathbb{R}^6$ and $L = \text{span}\{e_1, e_2, e_3, e_4\}$.

$$\dim(L \cap K) = 3: \quad K = \text{span}\{e_1, e_2, e_3, e_5, e_6\},$$
$$\dim(L \cap K) = 4: \quad K = \text{span}\{e_1, e_2, e_3, e_4, e_5\}.$$

(d) If $L^\perp \subseteq K$, then $L + K$ contains $L + L^\perp = \mathbb{R}^6$. Therefore

$$\boxed{\dim(L \cap K) = 4 + 5 - 6 = 3.}$$

What Can Be Determined from a Null Space?

QUESTION 11

Let $A, B \in \mathbb{R}^{m \times n}$.

(a) If $\mathcal{N}(A) = \mathcal{N}(B)$, must $\mathcal{C}(A^T) = \mathcal{C}(B^T)$? Prove or give a counterexample.

(b) If $\mathcal{C}(A) = \mathcal{C}(B)$, must $\mathcal{N}(A) = \mathcal{N}(B)$? Prove or give a counterexample.

(c) If $\mathcal{C}(A) = \mathcal{C}(B)$, prove that for every $b \in \mathbb{R}^m$,

$$\min_x \|Ax - b\|_2 = \min_y \|By - b\|_2.$$

(d) Does the assumption in (c) imply that A and B have the same rank? Justify.

What Can Be Determined from a Null Space?

ANSWER 11

(a) Yes. Use $\mathcal{C}(A^T) = \mathcal{N}(A)^\perp$:

$$\mathcal{N}(A) = \mathcal{N}(B) \Rightarrow \mathcal{C}(A^T) = \mathcal{C}(B^T).$$

(b) No. For example,

$$A = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \end{pmatrix}.$$

Both have column space $\mathbb{R}$, but distinct null spaces. **(c) Yes.** Each minimum is the distance from b to the same subspace $\mathcal{C}(A) = \mathcal{C}(B)$. **(d) Yes.** Since rank is column-space dimension,

$$\boxed{\text{rank}(A) = \dim \mathcal{C}(A) = \dim \mathcal{C}(B) = \text{rank}(B).}$$

QR Decomposition over $\mathbb{C}$

QUESTION 12

Let

$$A = \begin{pmatrix} 1 & i \\ i & 0 \\ 0 & 1 \end{pmatrix} \in \mathbb{C}^{3\times 2}, \qquad i^2 = -1.$$

(a) Use Gram–Schmidt to compute the reduced QR decomposition $A = QR$, with $Q^*Q = I_2$ and positive diagonal entries of R.

(b) Verify $Q^*Q = I_2$ and $QR = A$.

(c) For $b = (1, 0, 0)^T$, use QR to find

$$\widehat{x} = \operatorname*{argmin}_{x\in\mathbb{C}^2} \|Ax - b\|_2^2.$$

QR Decomposition over $\mathbb{C}$

ANSWER 12

(a) Gram–Schmidt gives

$$r_{11} = \sqrt{2}, \qquad r_{12} = q_1^* a_2 = \frac{i}{\sqrt{2}}, \qquad r_{22} = \sqrt{\frac{3}{2}}.$$

Hence

$$\boxed{Q = \begin{pmatrix} 1/\sqrt{2} & i/\sqrt{6} \\ i/\sqrt{2} & 1/\sqrt{6} \\ 0 & 2/\sqrt{6} \end{pmatrix}}, \qquad \boxed{R = \begin{pmatrix} \sqrt{2} & i/\sqrt{2} \\ 0 & \sqrt{3/2} \end{pmatrix}}.$$

(b) $q_1^* q_1 = q_2^* q_2 = 1$ and $q_1^* q_2 = 0$; therefore $Q^* Q = I_2$. Direct multiplication gives $QR = A$. **(c)** Solve $R\widehat{x} = Q^* b$:

$$\begin{pmatrix} \sqrt{2} & i/\sqrt{2} \\ 0 & \sqrt{3/2} \end{pmatrix} \widehat{x} = \begin{pmatrix} 1/\sqrt{2} \\ -i/\sqrt{6} \end{pmatrix} \quad \Longrightarrow \quad \boxed{\widehat{x} = \begin{pmatrix} 1/3 \\ -i/3 \end{pmatrix}}.$$