Element-Wise Application of Sigmoid Function on Full-Rank Square Matrices

Yuhao Liu

October 2023

1 Introduction

An empirical observation reveals that when the sigmoid function is applied element-wise to a full-rank matrix, the resulting matrix consistently maintains its full rank. This observation underscores that this transformation does not introduce linear dependencies between columns.

Moreover, bolstering this empirical observation, a compelling mathematical proof affirms that the resulting matrix is invariably full rank with a probability of 1.

2 Formal Verification

Definition 2.1. The sigmoid function, denoted as $\sigma(x)$, is a map from $\mathbb{R}$ to $(0, 1)$. It is defined as follows:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Lemma 2.1. *The sigmoid function is a bijection.*

Proof. The inverse of the sigmoid function is

$$\sigma(x)^{-1} = \ln(\frac{1 - x}{x})$$

□

Remark. The inverse of sigmoid function maps the $(0, 1)$ to $\mathbb{R}$.

Lemma 2.2. *The sigmoid function is a differentiable and the derivative is bounded.*

Proof. The derivative of the sigmoid function is $\sigma(x)(1 - \sigma(x))$, and clearly it is bounded since $\sigma(x)$ is bounded. □

Definition 2.2. A square matrix A is said to be of *full rank* if and only if the rank of matrix A, denoted as $\text{rank}(A)$, is equal to the size of the matrix, denoted as n. Mathematically, a square matrix A of size $n \times n$ is considered full rank if

$$\text{rank}(A) = n.$$

This condition implies that the rows and columns of matrix A are linearly independent, and the matrix has the maximum possible rank for its size.

Remark. It is well-known that when a random $n \times n$ square matrix is generated, the probability of it being full rank is 1. This phenomenon arises from the fact that the set of singular matrices forms a sub-manifold of dimension $2n - 1$. Consequently, the set of singular matrices has zero measure when considered within the context of Lebesgue measure.

Theorem 2.3. *Let A be a full rank $n \times n$ square matrix. Then, the element-wise application of the function σ to matrix A, denoted as $\sigma(A)$, is also full rank with probability* 1.

Proof. The sigmoid function applied element-wise to $n \times n$ matrices is a mapping from ^{2n} to $(0,1)^{2n}$. Since $\sigma(x)$ is differentiable with a bounded derivative (denote one of the bounds as c), we claim that the image of the set of singular matrices, denoted by N, under $\sigma(x)$ is a null set within $(0,1)^{2n}$.

Since N is a null set, there exists a collection $\{U_i\}$ of $2n$-dimensional cubes with a total volume of $\frac{\epsilon}{c^{2n}}$ that covers N. Therefore, the image of N is covered by the image of $\bigcup U_i$ with a total volume $\leq \epsilon$. Since ϵ is arbitrary, the image of N is a null set in $(0,1)^{2n}$.

The complement of N is the set of non-singular matrices (denoted by M), which will be mapped to the complement of a null set in $(0,1)^{2n}$. Note that the Lebesgue measure on $(0,1)^{2n}$ is indeed a probability measure. Hence the measure of M is 1.

Since the set of singular matrices in $(0,1)^{2n}$ is a null set (denoted by N'), the intersection of $\sigma(M)$ and N' is also a null set. Therefore, the probability of a non-singular matrix A being mapped to a singular matrix is 0. $\square$

Remark. In fact, a bijective function with bounded derivative applied element-wise to full-rank square matrices will be mapped to full-rank matrices with probability 1.