Interleaved Network Layers

Maria Walch

June 2022

1 Introduction

Let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ in the following be called the *input vector* and the *output vector*, respectively.

Definition 1.1 (Fully Connected Layer)**.** A fully connected layer is a function from $\mathbb{R}^n$ to $\mathbb{R}^m$. Each component of the output vector $\mathbf{y} \in \mathbb{R}^n$ depends on each component of the input vector $\mathbf{x} \in \mathbb{R}^m$.

Let $y_i \in \mathbb{R}$ be the i-th component of the output from the fully connected layer. Then $y_i \in \mathbb{R}$ is computed as follows:

$$y_i = a(w_1 x_1 + \dots w_n x_n), \tag{1}$$

where a denotes the nonlinear activation function. The $w_i \in \mathbb{R}$ are the learnable parameters of the network called the weights.
The full output is written as:

$$\mathbf{y} = \begin{bmatrix} a(w_{11}x_1 + \dots + w_{1n}x_n) \\ \vdots \\ a(w_{m1}x_1 + \dots w_{mn}x_n) \end{bmatrix} \tag{2}$$

This can also be written as:

$$\mathbf{y} = a(W \cdot \mathbf{x} + \mathbf{b}), \tag{3}$$

where $\mathbf{b} \in \mathbb{R}^m$ is called the bias vector. The weight matrix W with entries in $\mathbb{R}$ is given by:

$$W = \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{1m} & \dots & w_{mn} \end{bmatrix} \tag{4}$$

We first consider the operation $W\mathbf{x}$, which essentially gives a linear associator. The addition of $\mathbf{b}$ corresponds to a simple translation in the vector space $\mathbb{R}^m$.

2 The Singular Value Decomposition of W

The matrix W is a real $m \times n$ matrix and thus $W^T W$ is an $n \times n$ symmetric matrix with nonnegative eigenvalues ([For15], Lemma 7.4).

Definition 2.1 (Singular Values)**.** The singular values $\{\sigma_i\}$ of an $m \times n$ matrix A are the square roots of the eigenvalues of $A^T A$.

The singular value decomposition allows to factor a real $m \times n$ matrix A into a product of two orthogonal matrices and a diagonal matrix whose entries are the singular values of A.

Theorem 2.1 (Singular Value Decomposition ([For15], Theorem 15.1))**.** Let $A \in \mathbb{R}^{m \times n}$ be a matrix having r positive singular values, $m \leq n$. Then there exist orthogonal matrices $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$ and a diagonal matrix $\tilde{\Sigma} \in \mathbb{R}^{m \times n}$ such that

$$A = U \tilde{\Sigma} V^T \tag{5}$$

where

$$\tilde{\Sigma} = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \tag{6}$$

where $\Sigma = \text{diag}(\sigma_1, \sigma_2, ..., \sigma_r)$ and $\sigma_1 \geq \sigma_2 ... \geq \sigma_r > 0$ are the positive singular values of A.

Thus W can also be decomposed into the factorization $U \tilde{\Sigma} V^T$, consisting of the two orthogonal matrices U and V and the diagonal singular value matrix $\tilde{\Sigma}$. Intuitively,the orthogonal matrices U and V correspond to rotations or other transformations in $\mathbb{R}^m$ such as reflections, permutations or combinations thereof. The multiplication with $\tilde{\Sigma}$ corresponds to an anisotropic scaling.

If the kernel of the maps represented by U and V is zero, then the the underlying transformations are isometric.

Theorem 2.2 (Orthogonal Transformations and Isometries ([Mun91], Theorem 20.5))**.** Let $h : \mathbb{R}^n \to \mathbb{R}^n$ be a map such that $h(0) = 0$. Then the map h is an isometry if and only if it is an orthogonal transformation.

Thus multiplication of $\mathbf{x}$ by W corresponds to an orthogonal transformation of the entries x_i within $\mathbb{R}^n$ followed by a inhomogeneous dilation from $\mathbb{R}^n \to \mathbb{R}^m$. Within $\mathbb{R}^m$, there subsequently is another orthogonal transformation.

Theorem 2.3 (Rank and Singular Values ([For15], Theorem 15.3))**.** The rank of a matrix A is the number of nonzero singular values.

By the rank-nullity theorem, if the matrix A has full rank, then also $\tilde{\Sigma}$ has full rank.

3 Barcode of x

The entries of the input vector define a finite set of points X from a subspace $\mathbb{X} \in \mathbb{R}^n$. This is also called point cloud data or *PCD* for short.

Example 3.1 (PCD and Topological Invariants ([ZC05], Example 1.1))**.** If the sampling of $\mathbb{X}$ is dense enough, its topological invariants should be directly computable from the PCD.

References

[For15] William Ford. *Numerical Linear Algebra with Applications*. Oxford: Elsevier, 2015.

[Mun91] James R. Munkres. *Analysis on Manifolds*. Boca Raton: CRC Press, 1991.

[ZC05] Afra Zomorodian and Gunnar Carlsson. "Computing Persistent Homology". In: *Discrete and Computational Geometry* 33 (Feb. 2005), pp. 249–274. DOI: 10.1007/s00454-004-1146-y.