

Redes Neuronales 2020

FaMAF, UNC

Práctico 2

Nahuel Almeida

26 de noviembre de 2020

1. Introducción

En este práctico trabajaremos con una red neuronal de tipo *autoencoder* sobre el conjunto de datos MNIST.

Un autoencoder es un tipo de red neuronal *feed-forward*, cuyo objetivo es aprender la función identidad sobre un conjunto de datos. Una manera de representar esta red es la siguiente. Supongamos que contamos con un dataset $X \in \mathcal{R}^n$, cuyos elementos, que llamaremos *ejemplos* o *datos puntuales*, denominaremos como x . Un autoencoder cuenta con dos etapas. Una etapa de codificación, o *encoder*, donde se aplica una función f sobre los datos para producir el código $h = f(x)$. Luego, este código es decodificado en la segunda parte del autoencoder, denominada *decoder*, mediante una función g , tal que $\hat{x} = g(h) = g(f(x))$. En general, la salida de la red no coincide exactamente con el dato ingresado, sino que es una versión ligeramente distinta del mismo. La idea del autoencoder es aprender la función identidad de manera aproximada, y no de forma exacta.

Entre las aplicaciones que pueden tener los autoencoders mencionaremos dos. La primera tiene que ver con el código h . Si la dimensión de h es menor que la dimensión de x , entonces la red es capaz de reducir la dimensionalidad de los datos sin perder demasiada información. En este caso, no es la salida de la red lo que nos interesa, sino más bien el código en sí, puesto que nos está revelando información relevante de nuestros datos de manera comprimida.

La segunda aplicación son los autoencoders con corrección de ruido. En este caso, se busca que, al recibir una entrada \tilde{x} correspondiente a una versión ligeramente corrupta o modificada de x , el autoencoder sea capaz de revertir el cambio y devolver una versión corregida, más cercana al dato original.

El desempeño del autoencoder se puede medir con diferentes funciones de costo. En este práctico emplearemos el error cuadrático medio (MSE) definido como

$$L(\hat{x}, x) = \frac{1}{n} \sum_{i=1}^n (\hat{x}_i - x_i)^2. \quad (1)$$

Es decir, evaluaremos cada ejemplo teniendo en cuenta

en qué medida difiere (pixel por pixel) la salida de nuestro modelo del dato original.

2. Dataset

El dataset MNIST [1] consiste imágenes de dígitos escritos a mano, recolectados por el Instituto Nacional de Estándares y Tecnología de Estados Unidos (NIST, por sus siglas en inglés). Las imágenes tienen un tamaño de 28×28 píxeles en escala de grises con 256 valores de intensidad. Además, han sido preprocesadas de manera tal de que los números se ubiquen relativamente centrados y verticales. Cada imagen cuenta a su vez con una etiqueta, que informa el dígito correspondiente.

En la figura 1 se pueden ver algunos de los elementos del dataset. Podemos apreciar que, a simple vista, la mayoría de los ejemplos son fácilmente reconocibles para el ojo humano. Sin embargo, existen algunos casos donde existe cierta ambigüedad en el dígito, haciendo que el mismo sea difícil de reconocer por una persona no entrenada.

El dataset está dividido en un conjunto de entrenamiento de 60000 imágenes, y un conjunto de evaluación de 10000.

Referencias

- [1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE*, 86(11):2278-2324, November 1998.

١١١

١١١