

Red neuronal convolucional para la clasificación de imágenes

Martín Trucco¹

*Facultad de Matemática, Astronomía y Física - Universidad Nacional de Córdoba
Av. Medina Allende N°2144, Ciudad Universitaria X5016LAE, Córdoba, Argentina
martintrucco@mi.unc.edu.ar¹*

28 de Febrero, 2023

En el presente trabajo se empleó una red neuronal convolucional con dos capas convolucionales y tres capas intermedias sobre el conjunto de datos Fashion-MNIST. Para su diseño se seleccionaron diferentes hiperparámetros con el fin de encontrar aquellos que mejor ajusten al modelo. Se graficaron los resultados obtenidos para el error promedio de entrenamiento y testeo y la precisión de éstos en función de las épocas iteradas. Finalmente se comparó la red optimizada con una red feedforward de una sola capa oculta.

Descriptores: Redes neuronales convolucionales, CNN, ConvNets, redes neuronales.

In the present work, a convolutional neural network with two convolutional layers and three fully connected layers was used on the Fashion-MNIST dataset. For the design, different hyperparameters were selected in order to find those who fit the model better. Both the train and test average error and its precision were plotted as a function of the epochs. Finally, the optimized network was compared with a single hidden layer feedforward network.

Keywords: Convolutional neural networks, CNN, ConvNets, neural networks.

Introducción

Una red neuronal convolucional (CNN) es un algoritmo que toma un objeto de entrada *-input-*, a través del aprendizaje, asigna una importancia particular a ciertos aspectos o características del mismo que luego es capaz de diferenciar y filtrar. El preprocesamiento requerido en una CNN es mucho menor comparado con otros algoritmos de clasificación, lo que las convierte en herramientas sumamente útiles y con un campo de aplicación muy amplio como el procesamiento de imágenes y reconocimiento de voz [1], conducción de vehículos autónomos [2] o en radiología [3].

Estos modelos se inspiran en los patrones de conexión de las neuronas en el cerebro humano y la organización de la corteza visual. Cada neurona responde individualmente a un estímulo solo en una restringida región del campo visual conocido como el *campo receptivo* y una colección de estos campos se solapan para cubrir la totalidad del área visual [4].

En el presente trabajo se implementa una CNN que, mediante el estudio previo de su arquitectura, optimiza la clasificación de imágenes del dataset Fashion-MNIST. Posteriormente se comparan estos resultados con los de una red neuronal artificial (ANN) de una sola capa oculta con parámetros prefijados.

1 Marco Teórico

La arquitectura de las CNN puede simplificarse en tres elementos distintivos [5][6]: capas de convolución, capas de reducción de la muestra *-pooling-* y capas totalmente conectadas *-fully connected-*, que se esquematizan en la figura 1.

En el caso de una imagen, el elemento de entrada contiene los valores de los píxeles de la misma. La capa convolucional determina la salida de las neuronas que están conectadas a regiones locales del input a través del cálculo de un producto escalar entre estas regiones y unos filtros, también llamados *kernels*. Después de la convolución se aplica una función de activación (ReLU, sigmoide, tanh) para introducir la no linealidad en el algoritmo.

La capa de pooling tiene como objetivo reducir gradualmente la dimensionalidad de la representación y, por tanto, disminuir aún más la cantidad de parámetros, la capacidad computacional y la complejidad del modelo.

Por último la capa fully connected realiza una clasificación basada en las características extraídas de las capas anteriores. Por lo general, esta capa es una ANN tradicional con una función de activación que genera una probabilidad (un número que va de 0 a 1) para cada una de las etiquetas de clasificación que el modelo intenta predecir.

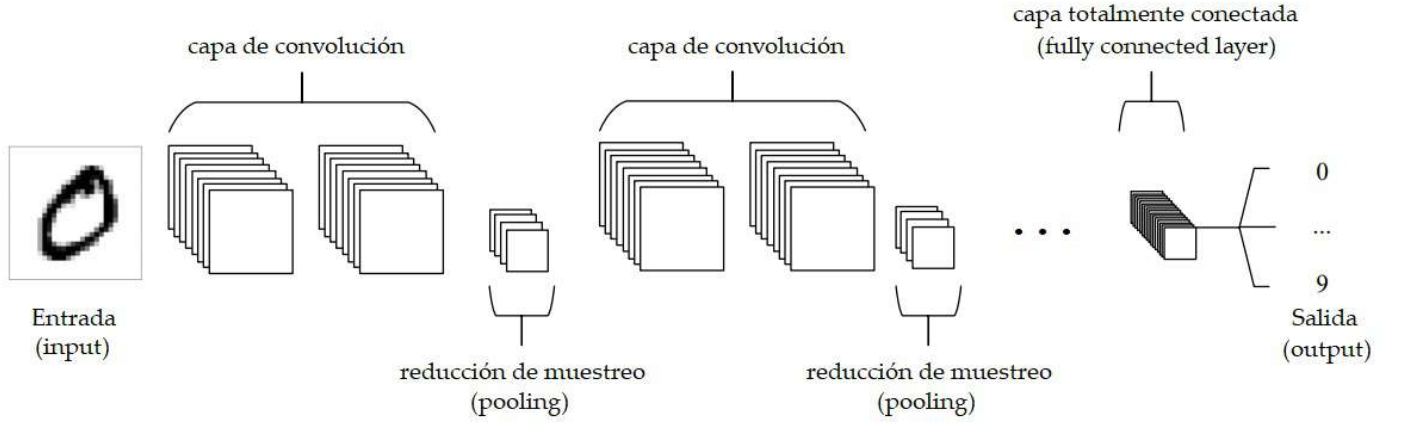


Figura 1: Arquitectura típica de una red neuronal convolucional. Ilustración adaptada de [5].

2 Procedimiento

Se crea una CNN con $28 \times 28 = 784$ unidades de entrada, 10 de salida, dos capas convolucionales de 32 y 64 canales respectivamente y tres capas intermedias (o densas) con una cantidad variable de neuronas n_N . Luego se emplea el dataset Fashion-MNIST con los conjuntos de entrenamiento y testeo predeterminados y se utiliza CrossEntropy como función de pérdida, una función de activación ReLU, filtros de tamaño 3×3 y Adam como algoritmo de descenso por el gradiente estocástico.

Se realiza un análisis de esta red para un número de épocas fijo $n_E = 5$ pero variando los valores de dropout d , ratio de aprendizaje l_r y n_N . Una vez elegidos estos tres hiperparámetros óptimos, se simula con $n_E = 50$ y se grafican los errores de entrenamiento y testeo en función de las épocas. De esta forma se determina el n_E necesario para maximizar el modelo.

Por último se repite el procedimiento con una

red ANN de parámetros ya establecidos para poder evaluar los errores, precisión y probabilidad de clasificación correcta entre ambos tipos de diseños.

3 Resultados y Discusión

3.1 Diseño de la red CNN

Definiendo $l_r = 0,001$, los valores de prueba n_N y d se exponen en la tabla 1 junto al porcentaje de precisión de testeo para cada uno de ellos.

El tamaño de la entrada en la primera capa densa se debe a la aplicación de un “achatao” -*flatten*- de tamaño 6×6 al canal de la segunda convolución, transformándolo en un vector de $64 \times 6 \times 6 = 2304$ elementos. Por otra parte, el tamaño de la última capa (de 10 neuronas) es la de salida.

También se elige $l_r = 0,01$ pero la precisión obtenida es del 10% para todos los d en cada época, indicando que no es una buena elección y se descarta.

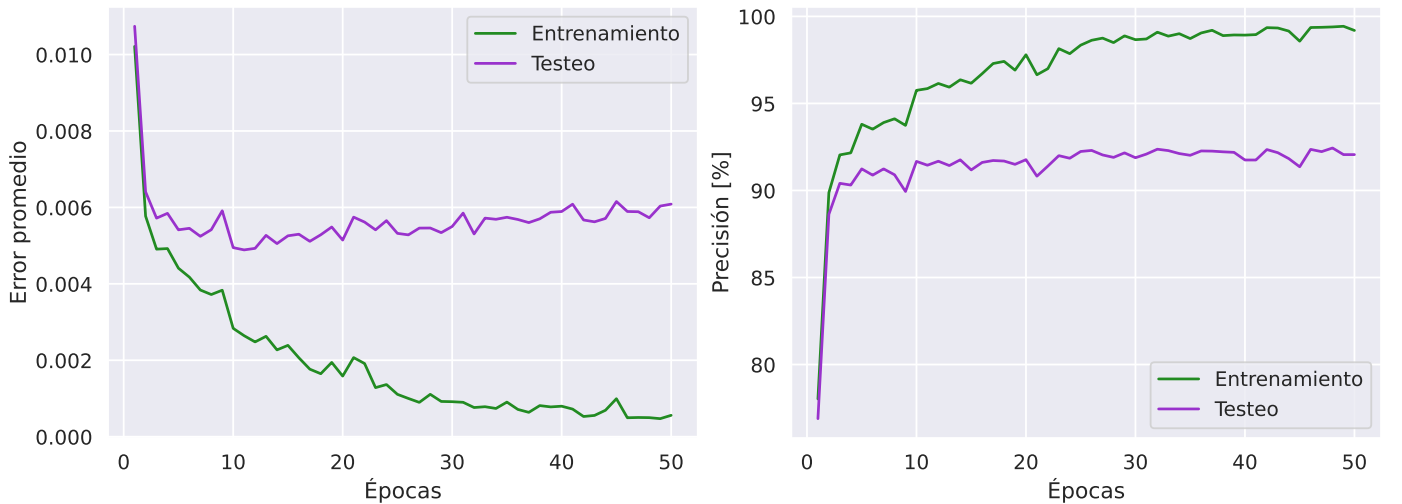


Figura 2: Errores promedio (izq.) y precisión (der.) de la CNN en función del número de épocas, $n_E = 50$.

n_N	d		
	0,1	0,25	0,4
2304–600–120–10	83,36	82,1	91,19
2304–600–300–10	79,13	49,00	90,76
2304–300–150–10	80,49	91,37	90,58

Tabla 1: Precisión de testeo (en [%]) para distintos dropouts d y números de neuronas en las capas densas n_N . Se resalta el porcentaje más alto alcanzado y el escogido para los análisis posteriores.

3.2 Comparación con la red ANN

La figura 2 exhibe el desempeño de la red con $n_E = 50$. A partir de la época 15 aproximadamente el error de testeo comienza a aumentar mientras que el de entrenamiento continúa disminuyendo, produciendo un sobreajuste. Por esta razón se considera $n_E = 15$ suficiente para estudiar el modelo. Teniendo en cuenta el trabajo de [7] se escogen para la ANN 64 neuronas en la capa oculta y un $d = 0,2$. Sin embargo, n_E , l_r , la función de activación, la función de pérdida y el optimizador son los mismos que para la CNN. Seguidamente se grafican los errores promedio y la precisión de ambos métodos optimizados en función de las épocas iteradas, los cuales se muestran en la figura 3.

Como se puede ver, los errores promedio son menores en la red convolucional y con una mayor precisión, alcanzando un porcentaje máximo de 91,19% en la época 9 en tanto que la red tradicional logra un pico máximo de 87,39% en la época 14.

Esta diferencia de clasificación se refleja en la tabla 2 donde, al tomar como input el elemento “T-Shirt” de la base de datos, la probabilidad de asignarle la etiqueta correcta al input difiere un 18% entre ambos.

Etiqueta	Probab ANN	Probab CNN
T-Shirt	0,81709	0,99889
Shirt	0,03289	0,00012
Ankle Boot	0,01875	0,00012
Dress	0,01875	0,00012
Bag	0,01875	0,00012
Trouser	0,01875	0,00012
Pullover	0,01875	0,00012
Sandal	0,01875	0,00012
Coat	0,01875	0,00012
Shirt	0,01875	0,00012
Sneaker	0,01875	0,00012

Tabla 2: Etiquetas con mayor probabilidad de ser asignadas al input “T-Shirt” según el modelo neuronal.

4 Conclusión

Se implementó satisfactoriamente una red neuronal convolucional a la base de datos del Fashion-MNIST mediante la selección de ciertos hiperparámetros analizados previamente, logrando clasificar imágenes con un 91,19% de precisión.

Se comparó con una red neuronal de una sola capa oculta, verificando su mejor desempeño como cabía esperar debido a un mayor número de parámetros. No obstante, ambos modelos tuvieron ratios de clasificación aceptables aunque es probable que las diferencias se acentúen si se utilizan inputs más complejos.

Finalmente se comprobó, dada la escasa cantidad de épocas necesarias para alcanzar el rendimiento obtenido, que el diseño de esta red es computacionalmente muy eficiente, lo que explica el extendido uso que tienen hoy en día.

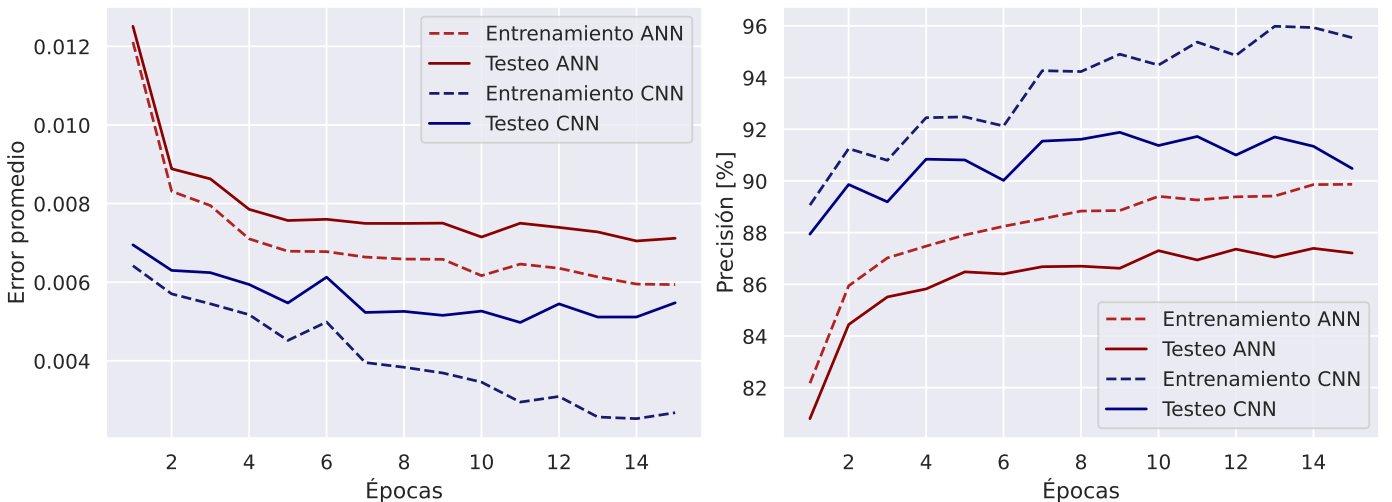


Figura 3: Errores promedio (izq.) y precisión (der.) de la ANN y CNN en función del número de épocas, $n_E = 15$.

Bibliografía

- [1] Alsobhani, A., ALabboodi, H. M. A., Mahdi, H. (2021). Speech Recognition using Convolution Deep Neural Networks. *Journal of Physics: Conference Series*, 1973(1), 012166. <https://doi.org/10.1088/1742-6596/1973/1/012166>.
- [2] Wu, Qisong Gao, Teng Lai, Zhichao Li, Dianze. (2020). Hybrid SVM-CNN classification technique for human-vehicle targets in an automotive LFM CW radar. *Sensors*. **20**. 3504. <https://doi.org/10.3390/s20123504>.
- [3] Yamashita, R., Nishio, M., Do, R.K.G. *et al.* (2018). Convolutional neural networks: an overview and application in radiology. *Insights Imaging* **9**, 611–629. <https://doi.org/10.1007/s13244-018-0639-9>.
- [4] Saha, S. (2022, November 16). A Comprehensive Guide to Convolutional Neural Networks - the ELI5 way. *Medium*. <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>
- [5] O'Shea, K. (2015, November 26). An Introduction to Convolutional Neural Networks. *arXiv.org*. <https://arxiv.org/abs/1511.08458>
- [6] *Deep Learning Series, P2: Understanding Convolutional Neural Networks*. (2019, January 26). Towards Machine Learning. <https://towardsml.wordpress.com/2018/10/16/deep-learning-series-p2-understanding-convolutional-neural-networks/>
- [7] Repositorio del curso de Redes Neuronales 2022, *Práctico 13*. FaMAF-UNC. GitHub - [jipphysics/redes-neuronales-2022](https://github.com/jipphysics/redes-neuronales-2022). GitHub. <https://github.com/jipphysics/redes-neuronales-2022>