Facultad de Ciencias Económicas

Universidad Nacional De Córdoba

Licenciatura en Economía

Trabajo Final Integrador:
Red neuronal profunda y convolucional para clasificar imágenes
Agosto 2022

Camila Chediack Ciminari
DNI: 41919762
mail: camila.chediack.c@mi.unc.edu.ar

Introducción

El comportamiento del cerebro humano se puede describir en pocas palabras a través del funcionamiento de las neuronas, las cuales son células especializadas en la transmición de información a través de estímulos en forma de cargas eléctricas. Si este estímulo sobrepasa un cierto umbral entonces este se dispara hacia el axon que se va a encargar de enviar dicho estímulo hacia las neuronas adyacentes. Esto no quiere decir que las neuronas siguientes vayan a recibir la señal de esa neurona, ocurre que no se encuentran conectadas entre sí a través del axón y las dendritas, sino que hay un espacio entre ellas denominado "espacio sináptico" en los que navegan estos neurotransmisores y van a ser "atrapados" por las dendritas de las neuronas que le siguen sólo si hay muchos de ellos. Esto quiere decir, que hay una cierta probabilidad de que las neuronas adyacentes recepcionen la información de la neurona que realizó el disparo.

Las redes neuronales son un algoritmo que tiene por objetivo simular este comportamiento del cerebro para resolver problemas complejos. Las mismas están compuestas por una capa de entrada de la información, $N-1$ capas intermedias y ocultas, seguido de la capa de salida que es el resultado obtenido. Matemáticamente se puede describir el modelo a través de la siguiente función:

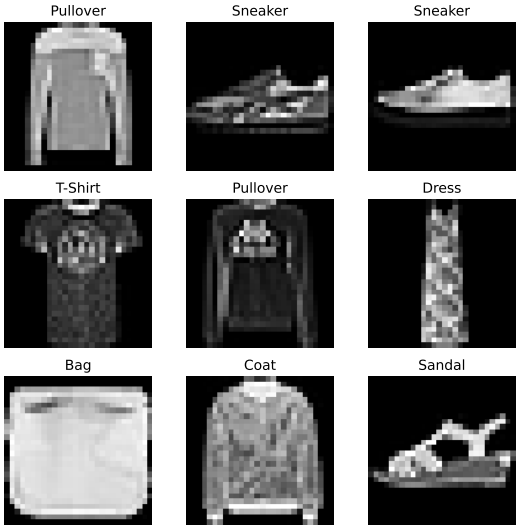
$$O_i^\mu = g(h_i) = g(\sum_k w_{ik}\xi_k)$$

Donde las relaciones entre las distintas neuronas se representan a través de la transmisión de información ponderada por los pesos sinápticos w_{ik}, el subíndice k hace alusión a la neurona de entrada e i a la neurona de salida. Si la información que le entra a la neurona sobrepasa el umbral, esta se dispara enviando la señal a las neuronas de la capa siguiente (si se trata de una red *feedforward*). Estos ponderadores dependen del modelo que se utilice y los hiperparámetros que se especifiquen, además de que se van actualizando a medida que se entrena la red, una característica puntual que tiene cualquier tipo de red neuronal es que esta va aprendiendo de los errores pasados con el objetivo de minimizar el mismo.

Red Neuronal clasificador de Fashion-MNIST

En esta ocasión se va a utilizar una Red Neuronal Profunda Convolucional para clasificar las distintas prendas de ropa que trae la base de datos de Fashion-MNIST. Las imágenes son de 28x28 píxeles en escala de grises, donde cada una de ellas se encuentra asociada a una de las 10 siguientes etiquetas: 0 T-shirt/top, 1 Trouser, 2 Pullover, 3 Dress, 4 Coat, 5 Sandal, 6 Shirt, 7 Sneaker, 8 Bag y 9 Ankle boot.

Imágenes de Fashion-MNIST



Fuente: Fashion-MNIST Dataset

La Red Neuronal Convolucional tiene la particularidad que las conexiones entre las distintas capas se realiza entre subconjuntos de neuronas. La forma en que estas recepcionan la información es muy parecida a las neuronas en la corteza visual primaria de un cerebro, y es por ello que se utilizan frecuentemente para clasificar o segmentar imágenes.

SU componente principal es la capa convolucional, la cual contiene distintos parámetros que establecen cómo va a ser la transformación de la imagen de entrada. Estos parámetros establecen la cantidad de filtros que se le van a aplicar a la misma, resultando en nuevas neuronas donde cada una de ellas se va a especializar en una sola característica propia de la imagen, obtienedo como resultado una nueva imagen con más dimensiones que se corre-

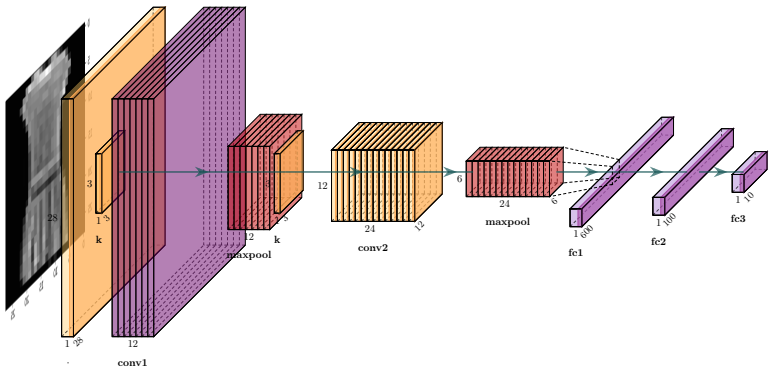
sponden con la cantidad de características en las que se descompuso la imagen original. Por cada capa de convolución van aumentando la cantidad de filtros (lo que es lo mismo decir que aumenta la cantidad de neuronas de salida) que luego van a ser concatenadas a través de una capa lineal, resultando en un super vector del tamaño de la cantidad de filtros por el alto y el ancho de cada uno de ellos.

Luego de obtener los resultados de su desempeño se va implementar a los mismos conjuntos de entrenamiento, validación y testeo, una red neuronal feed-forward sin convolución con una sola capa oculta para clasificar las distintas impágenes de prendas. El objetivo es realizar una comparación entre los desempeños de ambas redes, esperando obtener mejores resultados en la Red Neuronal Convolucional por todas las características mencionadas anteriormente sobre su funcionamiento.

Arquitectura de la Red Neuronal Convolucional

El input de la red es de $28x28x1$ por la cantidad de píxeles de la imagen y 1 dado que son en escala de grises. Se establecen dos capas de convolución con un *kernel* de $3x3$ y un *padding* igual a 1, seguida cada una de *Batch Normalization layer*, una función de activación *ReLU* y una *Max Pooling layer* con *kernel* de $2x2$ y un *stride* igual a 2. Luego siguen 3 capas lineales con un $dropout = 0.40$ para intentar evitar un *overfiting*.

Figure 1: **Red Neuronal Convolucional Profunda**



Fuente: Elaboración propia.

La operación que subyace de aplicarle 12 filtos con un *kernel* de altura=3,ancho=3 y pofundidad=1 es un producto punto entre esta matriz $3x3x1$ y distintas zonas de la imagen de entrada, dando como resultado una transformación de la misma con dimensión $28x28x12$ donde en vez de tener un solo filtro que correspondía al color de la imagen ahora se tienen 12 filtros de distintas características. Al establecer el parámetro *padding* igual a 1 esta operación también se aplica en los bordes y por lo tanto no hay una reducción de dimensionalidad. De la reducción de dimensionalidad se encarga la capa MaxPool con el objetivo de poder procesar las imágenes de forma más rápida.

Luego la primer capa lineal es la encargada de concatenar todos los filtros en un solo gran vector cuya entrada son 864 neuronas[1] y termina con 600 neuronas, seguido de una segunda capa con salida de 100 neuronas y finalmente la tercer capa lineal con 10 neuronas que corresponden a la cantidad de etiquetas que provee la base de datos Fashion-MNIST.

Usualmente en los modelos de clasificación se utiliza la función de pérdida *Cross-Entropy* dado que es más conveniente para medir la diferencia entre dos dsitribuciones de probabilidad de una variable dada, aumentando cuando la probabilidad de predicción de la variable diverge del label correcto.

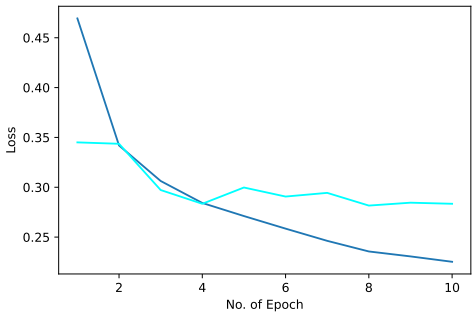
[1] Este número surge de multiplicar $6x6x24$ que se corresponde con el ouput de la segunda capa convolucional luego de ser aplicado el MaxPool

Como función de optimización se emplea la función Adam, vista como una combinación de las ventajas de AdaGrad y RMSprop, que utiliza una estimación del primer y segundo momento del gradiente (media y varianza) para adaptar la taza de aprendizaje para cada uno de los pesos de la red neuronal.

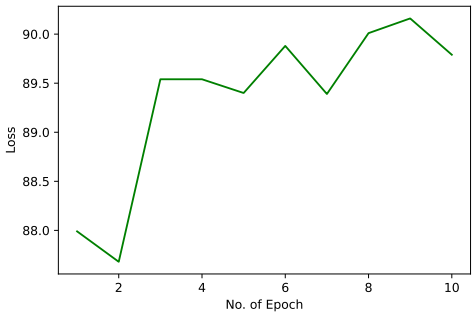
Resultados y comparación con una red sin convolución

Luego de entrenar los datos a través de la Red Neuronal Convolucional y validar los mismos durante 10 épocas se calcula la función de pérdida para el conjunto de entrenamiento y de validación, junto con la función de *accuracy* del modelo sobre los datos de validación para cada época.

Figure 2: Funciones de pérdida y precisión: Red Convolucional



(a) Training and Validation Loss



(b) Validation Accuracy

Fuente: Elaboración propia.

Como puede observarse en la figura 2, a partir de la época cuatro aproximadamente, se presenta un *overfitting*, esto quiere decir que la red sigue aprendiendo sobre los datos de entrenamiento pero cuando se prueba su desempeño en datos que desconoce, el error es mayor. Ajusta mejor los datos conocidos, con los que fue entrenada, que los datos de validación y posiblemente que los de testeo. Este resultado no es el deseable dado que el objetivo de entrenar la red es que extrapole bien para predecir sobre conjuntos de datos que nunca vió. Por otro lado la función de *accuracy* llega a su máximo en la época número 9 aproximadamente y en la 10 vuelve a bajar.

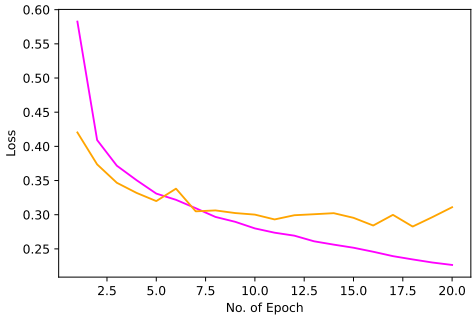
En la figura 3 la función de pérdida de validación es menor a la de entrenamiento solamente hasta las época 5 aproximadamento, para esa época en adelante también se produce un *overfitting*. En estos casos podría considerarse entrenar la red solamente hasta dicha época donde se nota un corte entre las dos funciones, para evitar que la red este aprendiendo de datos ruidosos que no aportan información de calidad.

Una posible respuesta a este *overfitting* es que las imágenes son muy simples para aplicar varias capas convolucionales con muchos filtros dada la poca cantidad de características que poseen en comparación a una imagen a color que contiene animales o rostros.

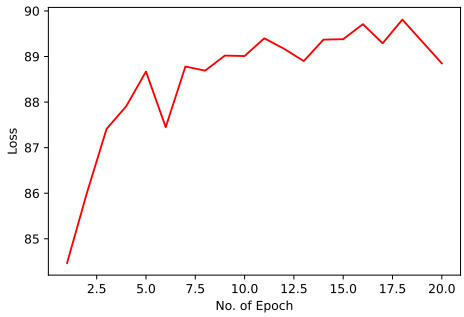
Otra razón que advierte este sobre entrenamiento son los buenos resultados que se obtuvieron del desempeño de la Red Neuronal con una sola capa oculta y sin convolución, los cuales se muestran a continuación junto con una tabla de comparación de ambos modelos a través de distintas métricas.

Una forma adecuada para poder comparar distintos modelos es a través del resultado de distintas métricas como precision, recall y f1 score que nos brinda el "Classification Report" de la librería de **sklearn**. Mientras que la primera nos dice el porcentaje de predicciones positivas correctas sobre el total de predicciones positivas, la recall es el porcentaje de predicciones positivas correctas sobre el total de positivos que hay y por último f1 score es un promedio ponderado entre las dos anteriores.

Figure 3: Funciones de pérdida y precisión: Red Neuronal sin convolución



(a) Training and Validation Loss



(b) Validation Accuracy

Fuente: Elaboración propia.

Table 1: **Reporte de clasificación: Red Neuronal Convolucional**

	Precision	**Recall**	**F1-score**
Accuracy	0.897800	0.8978	0.897800
macro avg	0.903567	0.8978	0.899302
weighted avg	0.903567	0.8978	0.899302

Table 2: **Reporte de clasificación: Red Neuronal sin convolución**

	Precision	**Recall**	**F1-score**
Accuracy	0.8799	0.8799	0.8799
macro avg	0.882789	0.8799	0.878057
weighted avg	0.882789	0.8799	0.878057

En la tabla 1 se muestran las métricas correspondientes a la red neuronal con la capa de convolución, la cual aproximadamente en el 90% de los casos predice de forma correcta la etiqueta de la imagen. En la tabla 2 se exponen los resultados para la red neuronal sin convolución, la cual muestra una precisión del 88% en todas las métricas, es decir que predice en forma correcta el 88% de las veces. La diferencia entre estos dos modelos es bastante sensible.

Conclusiones

En un comienzo se expresó la idea de que una red convolucional es mucho mejor para clasificar imágenes dada las características propias de este tipo de redes, sin embargo los resultados han demostrado una muy buena predicción por parte del modelo que no tiene capa de convolución y que solo tiene una capa oculta. Por lo tanto esto nos da un pauta para creer que para imágenes tan simples como las que trae el dataset de Fashion-MNIST bastaría con implementar una red sin convolución. A su vez se mostró que se da un overfitting en ambos casos, que en la primer red tal vez podría ser revertido aplicando solo una capa de convolución, sin embargo dejaría de ser una red neuronal convolucional profunda.

En cualquiera de los dos casos y a pesar de que se produce un *overfitting*, se obtienen buenos resultados con un alto porcentaje de predicciones correctas que seguramente podría ser mejorado si se simplifica la arquitectura de la red para evitar este sobre ajuste y que consiga extrapolar mejor, siendo el resultado deseable.