Universidad de Granada

Inteligencia Computacional

Máster en Ingeniería Informática

Reconocimiento óptico de números manuscritos en la base de datos MNIST con redes neuronales

Autor
Pedro Manuel Gómez-Portillo López

Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación

Granada, 27 de noviembre de 2018

Reconocimiento óptico de números manuscritos en la base de datos MNIST con redes neuronales

Pedro Manuel Gómez-Portillo López

Resumen

Las redes neuronales son una herramienta muy potente que pueden ser entrenadas para resolver todo tipo de problemas.

En esta práctica se trabajará con ellas para reconocer caracteres escritos a mano. Haciendo uso de la base de datos MNIST de números manuscritos y del framework de desarrollo Keras, se configurará y entrenará una red neuronal para que, al evaluarla, reconozca el mayor número de dígitos posibles.

Palabras clave: *MNIST, reconocimiento óptimo de caracteres, deep learning, keras*

El formato de la documentación de este trabajo ha sido basado en la plantilla LaTeX de https://github.com/erseco.

Índice general

Capítulo 1

Introducción

Las redes neuronales son un modelo computacional basado en un gran conjunto de neuronas individuales de forma aproximadamente análoga al comportamiento observado en los axones de las neuronas en los cerebros biológicos.

Actualmente existe una gran cantidad de programas que hacen uso de las redes neuronales para toda clase de aplicaciones; desde traducción de texto a clustering de datos.

Una de estas aplicaciones es el reconocimiento óptico de elementos. Concretamente, la base de datos MNIST[1] es un conjunto de imágenes con números escritos a mano. Esta base de dato tiene un conjunto de 60,000 imágenes, con un tamaño estándar de 28*28 píxeles cada una, y es la que se utilizará en esta práctica. En la figura 1.1 puede verse un ejemplo de los números almacenados en esta base de datos.

Figura 1.1: Ejemplo de números en la base de datos MNIST

Concretamente, se utilizará esta base de datos para entrenar una red neuronal y aplicarla al reconocimiento de estos caracteres, intentado acertar en el mayor número posible.

En esta práctica se podría optar o bien por desarrollar una red neuronal desde 0 o bien por utilizar un framework o conjunto de librerías ya desarrollado. Esta decisión se basará y se justificará en el siguiente capítulo.

[1] `http://yann.lecun.com/exdb/mnist/`

1.1. Entorno de desarrollo

Mi ordenador personal, en el cual se ha realizado esta práctica, cuenta con las siguientes características.

- **Sistema operativo**. Ubuntu 18.04.1 LTS (Bionic Beaver)
- **Procesador**. Intel Core i7-6700HQ Quad-Core 2.6GHz
- **Cache**. 6M
- **Memoria RAM**. 8GB DDR4 2,133MHz
- **Tarjeta gráfica**. Sí, pero no recodida por el sistema operativo, por lo que no podrá usarse.
- **Disco duro**. SanDisk SSD Plus 256GB

Capítulo 2

Estado del arte

Antes de elegir lenguaje de programación a usar se realizó una búsqueda de los frameworks para desarrollo de aplicaciones basadas en deep learning disponibles, y de este modo tener razones para poder justificar esta elección.

A continuación se presentan los frameworks más importantes que hay en el mercado.

2.1. Keras

Keras[1] es una librería de código abierto escrita en Python, diseñada específicamente para hacer experimentos con redes neuronales.

Lo interesante de este framework es que funciona como **front-end** de la librería que utilice por debajo, es decir, que redirige las llamadas a sus funciones a las de la librería con la que le digamos que trabaje. Keras puede ejecutarse sobre MXNet, DL4J, TensorFlow, o Theano, algunas de las que hablaremos más adelante.

2.2. Pytorch

Pytorch[2] es un framework para Python que permite el prototipado y desarrollo rápido de aplicaciones deep learning, especialmente centrado en la aceleración por GPU.

La característica principal de Pytorch es que utiliza grafos computacionales dinámicos por cuestiones de eficiencia y optimización, ya que según su API estos grafos se paralelizan especialmente bien en una tarjeta gráfica.

[1] `https://keras.io/`
[2] `https://pytorch.org/`

2.3. TensorFlow

TensorFlow[3] es es un framework de código abierto desarrollado por Google que, según su página web, *se utiliza para realizar cálculos numéricos mediante diagramas de flujo de datos; los nodos de los diagramas representan operaciones matemáticas y las aristas reflejan las matrices de datos multidimensionales (tensores) comunicadas entre ellas.*

Este framework trabaja a un nivel más bajo que Keras o Pytorch y, aunque es uno de los más populares y usados por empresas como Dropbox, su uso parece ser recomendado para proyectos más grandes y complejos que éste.

2.4. Scikit-learn

Scikit-learn[4] es un framework de código abierto escrita en Python utilizada por empresas como Spotify que, según su página web, es *simple, eficiente y accesible, perfecta para para técnicas de análisis y minería de datos.* Está escrita sobre otras librerías de Python como *NumPy*, *SciPy*, y *matplotlib*.

2.5. Theano

Theano[5] es un frameworks de bajo nivel, como TensorFlow, y es otro de los soportados por Keras. Según su página web permite *definir, optimizar y evaluar expresiones matemáticas, especialmente las que trabajan con matrices multi-dimensionales.*

Aún así, Theano no está especialmente centrado en el deep learning, por lo que en lugar de usar esta librería tiene más sentido usar frameworks como Keras que se apoyen en el para realizar tareas de *deep learning*.

2.6. Lasagne

Lasagne[6] es una librería escrita en Python que nació exclusivamente para permitir usar Theano en tareas de deep learning, y proporcionando una interfaz más amigable. Es una de las principales competidores de Keras, aunque parece no ser tan preferida como esta.

[3] `https://www.tensorflow.org/`
[4] `https://scikit-learn.org/`
[5] `http://www.deeplearning.net/software/theano/`
[6] `https://lasagne.readthedocs.io/`

2.7. DSSTINE

DSSTINE[7] (pronunciado *destiny*), es un framework de código abierto desarrollado por Amazon especialmente pensado para entrenar y desplegar modelos de recomendación.

Además, únicamente permite ejecutarse sobre GPU, aunque permite hacerlo en paralelo usando varias. Por otro lado, su documentación es muy pobre y a veces es necesario bucear en su código fuente para entender qué hace.

2.8. MXNet

MXNet[8] es una librería de uso general desarrollada por Apache. A pesar de asegurarse bastante potente parece estar en una fase muy verde.

2.9. DL4J

DeepLearning4J[9] es una librería distribuida de código abierto escrita en Java y disponible para Java, Python y C++.

Según su página web, se especializa en redes neuronales profundas, y su documentación parece muy completa y está muy bien escrita.

2.10. Microsoft Cognitive Toolkit

CNTK[10] es una librería de código abierto desarrollada por Microsoft Research, la división de investigación de Microsoft que, según su página web, *entrena algoritmos de deep learning para pensar como personas.*

A pesar de lo prometedor que parecía, no parece ser muy popular, ya que en comparación con el resto de frameworks no hay muchos ejemplos por la web, ni siquiera en páginas especializadas como Kaggle[11].

[7] `https://github.com/amzn/amazon-dsstne`
[8] `https://mxnet.incubator.apache.org/`
[9] `https://deeplearning4j.org/`
[10] `https://www.microsoft.com/en-us/cognitive-toolkit/`
[11] `https://www.kaggle.com/`

Capítulo 3

Implementación

Tras realizar el estudio acerca del estado del arte, y basándonos que necesitamos un framework centrado en deep learning que no necesite GPU para ejecutarse, se ha decidido utilizar el lenguaje de programación Python3 con el framework Keras con TensorFlow como back-end.

Cabe destacar que, aunque me hubiera parecido mucho más interesante programar la práctica desde 0, como se proponía en el guión, en lugar de utilizar una librería ya implementada, las restricciones de tiempo y esfuerzo a las que ha tenido que enmarcase esta práctica debido al resto de asignaturas han hecho que finalmente opte por utilizar una framework de desarrollo de aplicaciones *deep learning*.

3.1. Framework y librerías utilizadas

A continuación se presentan los programas más importantes utilizados en esta práctica, así como su versión específica. Se han obviado las dependencias.

- **Python 3.6.6**. Será el lenguaje de programación de la práctica.
- **Keras 2.2.4**. Se utilizará como front-end para trabajar con TensorFlow, ya que actúa como interfaz facilitando su uso.
- **TensorFlow 1.13.0**. Será el corazón de la aplicación
- **Matplotlib 3.0.0**. e utilizará para visualizar el contenido de la base de datos MNIST.

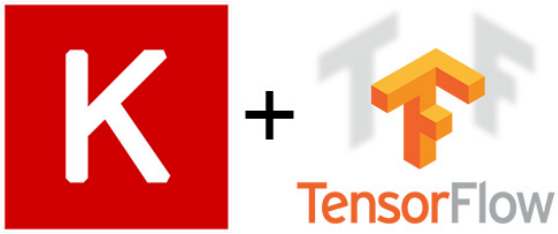

Figura 3.1: Logos de Keras y TensorFlow

Para instalar estas librerías usando el proyecto solo es necesario situarse en el directorio raíz y ejecutar `make install`.

3.2. Detalles comunes de la implementación

A continuación se presentan la información de la implementación del proyecto. Primero se presentará la información común a todas las versiones del proyecto y a continuación se pasará a describir cada versión por separado. Los resultados detallados de cada versión pueden verse en el capítulo 5, mientras que aquí solo se presentará la configuración de cada una.

En total, este proyecto ha pasado por tres versiones, pero es necesario destacar que las versiones no son totalmente sucesivas, sino que entre ellas se realizaron varias subversiones probando diferentes configuraciones de épocas y capas, y no se decidía definir una nueva versión hasta obtener una mejoría suficiente.

Además, se ha utilizado un repositorio en GitHub para realizar el control de versiones de la práctica, cuya dirección es `https://github.com/gomezportillo/mnist`.

Al final de cada versión se ha publicado una release del proyecto, por lo que las tres versiones pueden verse en la correspondiente sección del repositorio.

Todas las versiones utilizan las mismas librerías, por lo que para probar cualquiera de ellas, y habiendo clonado previamente el repositorio e instalado las dependencias, bastaría con situarse en el directorio raiz y ejecutar `git checkout vX.0`, donde `x` es la versión del proyecto, entre 1 y 3, a la que nos queremos mover, y escribir `make` para ejecutar la práctica en dicha versión.

3.3. Primera versión

La primera versión fue la más larga y complicada, ya que hubo que configurar todo el entorno de desarrollo para poder trabajar con las librerías elegidas.

En un esfuerzo por hacer los resultados de este proyecto reproducibles, se ha configurado un archivo `Makefile` con las instrucciones de instalación necesarias para volver a preparar el mismo entorno de desarrollo. Para instalarlo, simplemente es necesario situarse en el directorio raíz del proyecto y ejecutar `make install`, lo que básicamente instalará Python3, luego pip3 y lo usará para instalar las librerías indicadas en el archivo `requirements.txt`.

Tras instalar Keras y TensorFlow y sus dependencias fue posible empezar a trabajar en la práctica. Como nunca había trabajado con estas librerías acudí al sitio web de Keras y seguí su tutorial de iniciación[1]. En este tutorial se presenta el lenguaje y los principales pasos que hay que seguir para definir y configurar una red neuronal desde cero, así que aunque en el tutorial se utiliza la base de datos *Fashion MNIST*[2] de prendas de vestir, fue fácil adaptarla para utilizar la base de datos MNIST.

Tras haber configurado el entorno de desarrollo, haber entendido cómo trabajar con el framework Keras y haber obtenido mis primeros resultados, la configuración final de capas fue la siguiente, usando un modelo secuencial con 15 épocas de entrenamiento.

- La primera capa, llamada *Flatten*, transforma las imágenes de la base de datos de una matriz bidimensional de 28x28 píxeles a un vector unidimensional de 784 píxeles (28*28), lo que permite que la salida sea procesada por capas totalmente conectadas.
- La segunda capa, llamada *Dense*, es una capa totalmente conectada con 128 neuronas y una función de activación de rectificación lineal.
- La tercera capa, llamada *Dense*, es la capa de salida, con 10 neuronas (una por cada clase) y una función de activación *softmax* para poder realizar predicciones probabilísticas más tarde.

La imagen 3.2, generada con la función de Keras `plot_model` a partir del modelo, refleja la configuración de capas utilizada.

[1] `https://www.tensorflow.org/tutorials/keras/basic_classification`
[2] `https://github.com/zalandoresearch/fashion-mnist`

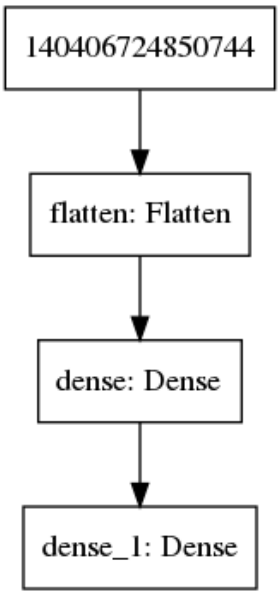

Figura 3.2: Configuración de capas de la primera versión

Con esta configuración tan simple pude obtener un porcentaje de acierto sobre el conjunto de entrenamiento del **99.8 %** y un **97.62 %** sobre el conjunto de prueba, que me pareció un resultado más que aceptable para una primera versión. Los resultados de esta versión pueden verse detallados en profundidad en el siguiente capítulo.

La ejecución de esta versión duró un poco más que la primera (300 segundos frente a los solamente 20 segundos de la primera), así que aunque al principio me planteé usar algún servicio PaaS[3] como *Amazon Web Services* para ejecutar la práctica, finalmente decidí que debido a su coste y a que los tiempos no eran excesivamente altos podía seguir ejecutándola en mi ordenador personal.

El enlace de esta versión en el repositorio es `https://github.com/gomezportillo/mnist/tree/v1.0`.

3.4. Segunda versión

Como ya tenía configurado el entorno de desarrollo y sabía cómo trabajar con el framework elegido, en la segunda versión pude navegar tranquilamente por la API de Keras[4] para conocer y entender todas las funciones y parámetros que tiene.

[3] Platform as a Service
[4] `https://keras.io/`

Al final realicé una configuración de 5 capas, usando un modelo secuencial con 15 épocas de entrenamiento.

- La primera capa, llamada *Conv2D*, es una capa convolutiva con la función de activación de rectificación lineal y una ventana convolutiva de 3x3.

- La segunda capa, llamada *MaxPooling2D*, es una capa de operación de agrupación máxima con un tamaño de pool de 2x2.

- La tercera capa, llamada *Dropout*, descarta el 25 % de las neuronas aleatoriamente para evitar problemas de sobreapendizaje en la fase de entrenamiento.

- La cuarta capa, llamada *Flatten*, convierte los datos de entrada de una matriz bidimensional de 28x28 a un vector unidimensional para que puedan ser procesados por capas totalmente conectadas.

- La quinta capa, llamada *Dropout*, tiene un número de nodos igual al número de clases del modelo, del 0 al 9, con la función de activación *softmax* para poder realizar predicciones más adelante.

La imagen 3.3 refleja la configuración de capas utilizada.

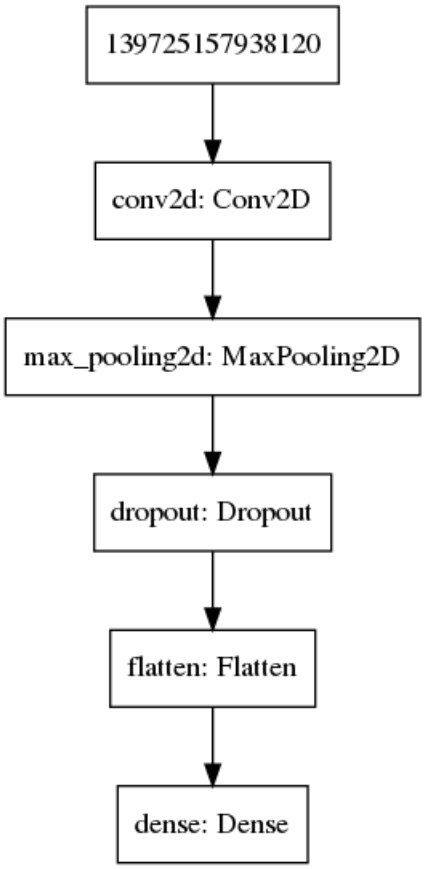

Figura 3.3: Configuración de capas de la segunda versión

Con esta configuración, algo más compleja, ya que usa capas convolutivas e intenta evitar el sobreaprendizaje de la red usando funciones dropout, fue posible obtener un porcentaje de acierto sobre el conjunto de entrenamiento de un **99.27 %** y un **98.22 %** sobre el conjunto de prueba.

Me pareció interesante que, comparado con la versión anterior, normalmente obtenía resultados algo peores en el conjunto de entrenamiento pero casi un 1 % mejores en el conjunto de prueba, lo que achaqué a que había solucionado los problemas de sobreaprendizaje.

El enlace de esta versión en el repositorio es `https://github.com/gomezportillo/mnist/tree/v2.0`.

3.5. Tercera versión

Por último, en la tercera versión del proyecto necesitaba un porcentaje de aciertos lo más cercano al 100 % posible tanto en el conjunto de entrenamiento como en el prueba.

El resultado final fue una configuración de 9 capas, usando un modelo secuencial con 20 épocas de entrenamiento. Como ya había resuelto el problema del sobreaprendizaje pude usar más épocas de entrenamiento que en las versiones anteriores.

- La primera capa es una capa convolutiva llamada *Conv2D* con una función de activación de rectificación lineal, 32 mapas característicos y una ventana convolutiva de 4x4.
- La segunda capa es una capa convolutiva llamada *Conv2D* con la función de activación de rectificación lineal y una ventana convolutiva de 3x3.
- La tercera capa, llamada *MaxPooling2D*, es una capa de operación de agrupación máxima con un tamaño de pool de 2x2.
- La cuarta capa, llamada *Dropout*, descarta el 20 % de las neuronas aleatoriamente para evitar problemas de sobreapendizaje en la fase de entrenamiento.
- La quinta capa, llamada *Flatten*, convierte los datos de la matriz bidimensional en un vector lineal para que la sexta capa pueda ser una capa completamente conectada.
- La sexta capa, llamada *Dense*, es una capa completamente conectada con 248 neuronas y función de activación de rectificación lineal.
- La séptima capa, llamada *Dense*, es una capa completamente conectada con 124 neuronas (la mitad que la anterior) y función de activación de rectificación lineal.
- La octava capa, llamada *Dropout*, vuelve a descartar el 40 % de las neuronas aleatoriamente.
- La novena capa, llamada *Dense*, es la capa de salida, con 10 neuronas (una por cada clase) y una función de activación *softmax* para poder realizar predicciones probabilísticas más adelante.

La imagen 3.4 refleja la configuración final de capas utilizada.

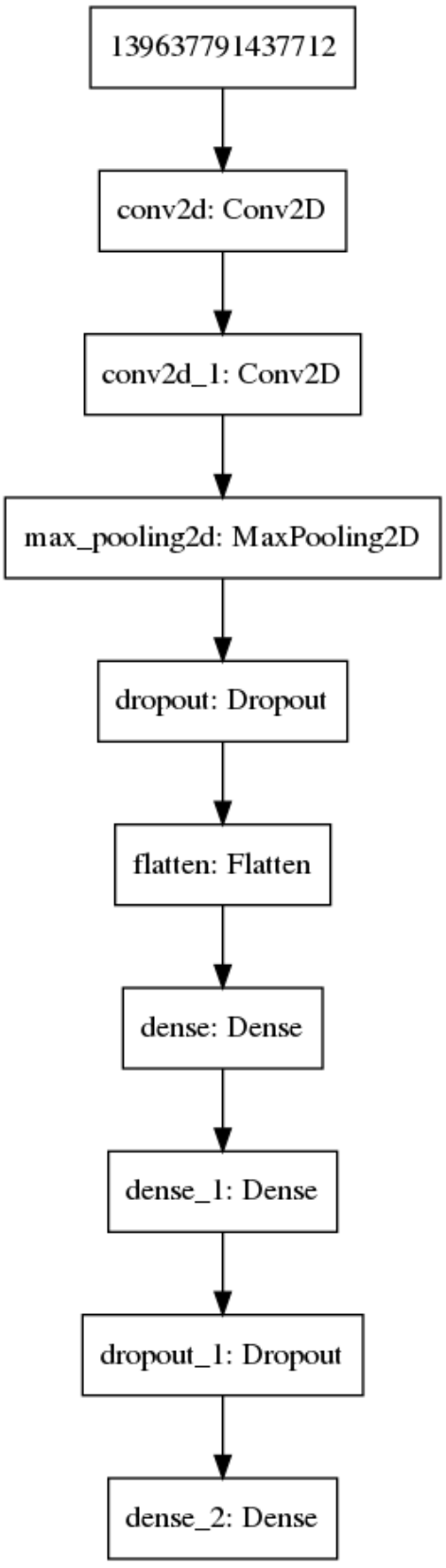

Figura 3.4: Configuración de capas de la tercera versión

Con esta configuración pude obtener un porcentaje de aciertos sobre el conjunto de prueba de un **99.91 %** y un **99.15 %** sobre el conjunto de prueba.

El enlace de esta versión en el repositorio es `https://github.com/gomezportillo/mnist/tree/v3.0`.

Capítulo 4

Resultados

A continuación se presentan los resultados detallados de las diferentes versiones del proyecto.

4.1. Primera versión

Tras ejecutar la primera versión, cuya fase de entrenamiento duró **22.2 segundos**, obtuve los siguientes resultados, formateados en la tabla 4.1. La columna *Época* indica a qué época del entrenamiento hacen referencia los resultados, la columna *Loss* es un valor escalar que intentamos reducir a 0 y la columna *Acierto* indica el porcentaje de aciertos que se han obtenido.

1whitelightgray

Época	Loss	Acierto	Época	Loss	Acierto
1	0.3603	0.9007	11	0.0271	0.9927
2	0.1674	0.9526	12	0.0240	0.9939
3	0.1205	0.9656	13	0.0209	0.9946
4	0.0929	0.9736	14	0.0174	0.9958
5	0.0750	0.9784	15	0.0149	0.9965
6	0.0622	0.9819			
7	0.0524	0.9853			
8	0.0439	0.9877			
9	0.0375	0.9895			
10	0.0326	0.9911			

Capítulo 5

Conclusiones

Capítulo 6

Referencias

Capítulo 7

Anexos