

Aspectos a considerar en Proyectos de Deep Learning

Juan Gamarra Moreno

Introducción

Generalmente los proyectos desarrollados con Keras implicarán actividades relacionadas a:

- Cargar datos (por ejemplo desde un archivo CSV) y ponerlos a disposición de Keras.
- Preparar datos de clasificación multiclase para modelar con redes neuronales.
- Definir los modelos de Deep Learning con Keras
- Evaluar los modelos de Deep Learning de Keras con scikit-learn.

Ejemplo de Proyecto de Clasificación Multiclase

- Entender el dataset
- Cargar las librerías necesarias
- Cargar el dataset
- Codificar la variable de salida
- Definir el modelo de Red Neuronal
- Evaluar el modelo con validación cruzada K-Fold

Ejemplo de Proyecto de Clasificación Binaria

- Entender el dataset
- Definición y evaluación del modelo base
- Mejora del desempeño con Preparación de Datos
- Ajuste de capas y neuronas en el modelo
 - Evaluar una red más pequeña
 - Evaluar una red más grande

Preparación de datos

- Es una buena práctica preparar sus datos antes de modelar. Los modelos de redes neuronales son especialmente adecuados para tener valores de entrada consistentes, tanto en escala como en distribución.

Preparación de datos

- Un esquema de preparación de datos efectivo para datos tabulares al construir modelos de redes neuronales es la estandarización. Aquí es donde los datos se reescalan de manera que el valor medio para cada atributo sea 0 y la desviación estándar sea 1. Esto preserva las distribuciones gaussianas y similares a las de Gauss mientras normaliza las tendencias centrales para cada atributo.

Preparación de datos

- En lugar de realizar la estandarización en todo el conjunto de datos, es una buena práctica entrenar el procedimiento de estandarización en los datos de entrenamiento dentro del pase de una ejecución de validación cruzada y usar la instancia de estandarización entrenada para preparar el fold de prueba no visto. Esto hace que la estandarización sea un paso en la preparación del modelo en el proceso de validación cruzada y evita que el algoritmo tenga conocimiento de datos no vistos durante la evaluación, conocimiento que podría ser pasado del esquema de preparación de datos como una distribución más nítida.
- Podemos lograr esto en scikit-learn usando una clase Pipeline. La canalización (pipeline) es un envoltorio que ejecuta uno o más modelos dentro de un paso del procedimiento de validación cruzada.

Ajuste de capas y neuronas en el modelo

- Hay muchas cosas para ajustar en una red neuronal, como la inicialización del peso, las funciones de activación, el procedimiento de optimización, etc.
- Un aspecto que puede tener un efecto descomunal es la estructura de la propia red denominada topología de red. Por ejemplo se podría experimentar sobre la estructura de la red al hacerla más pequeña y luego más grande. Estos son buenos experimentos a realizar al ajustar la red neuronal a su problema.

Ejemplo de Proyecto de Regresión

- Entender el dataset
- Definición y evaluación del modelo base
- Mejora del desempeño con Preparación de Datos
- Ajuste la topología de la red
 - Evaluar una topología de red más profunda
 - Evaluar una topología de red más ancha