

Técnicas de Aprendizaje Profundo aplicadas a la industria de retail.

Presentado por:
Juan Felipe Osorio López | Aura María Molina Amaya
Yackeline Cristina Quintero López | Juan José Toro Villegas

Planteamiento del problema

En el contexto empresarial de un servicio de reconocimiento de imágenes, nos enfrentamos al desafío de desarrollar un sistema de detección de objetos en imágenes utilizando el conjunto de datos CIFAR-10 proporcionado por Keras. Este problema tiene aplicaciones prácticas en diversas industrias, como la automotriz, la de seguridad, la de retail y la de atención médica.

Imaginemos una empresa dedicada a la gestión de inventarios en el sector retail. En este caso, el problema de negocio es mejorar la precisión y eficiencia del sistema de monitoreo de inventarios mediante la identificación y clasificación automática de productos en las estanterías. Esto es crucial para optimizar la reposición de stock, reducir pérdidas por errores humanos y mejorar la experiencia de compra de los clientes.

Diseño de la solución propuesta:

Definición del Modelo: Proponemos crear un modelo predictivo utilizando redes neuronales convolucionales (CNN) para identificar objetos en imágenes. Las CNN son particularmente eficaces en tareas de visión por computadora debido a su capacidad para detectar y aprender patrones complejos en datos visuales.

Arquitectura del Modelo: La estructura de la CNN se diseñará para extraer características distintivas de los objetos en las imágenes. Emplearemos capas convolucionales para la extracción de características y capas densamente conectadas para el aprendizaje de patrones jerárquicos. La configuración del modelo será determinada según la complejidad del problema y la cantidad de datos disponibles.

Preprocesamiento de Datos: Antes de alimentar los datos al modelo, realizaremos un preprocesamiento completo. Esto incluirá la normalización de los valores de los píxeles, técnicas de aumento de datos para evitar el sobreajuste y la segmentación de objetos para proporcionar información contextual adicional.

Conjunto de Datos: Utilizaremos el conjunto de datos CIFAR-10 para entrenar y validar el modelo. Este conjunto consta de 60,000 imágenes en color de 32x32 píxeles, divididas en 10 categorías diferentes de objetos, lo cual es adecuado para nuestro problema de detección de objetos.

Entrenamiento y Validación del Modelo: Entrenaremos el modelo usando el conjunto de datos CIFAR-10, dividiéndolo en subconjuntos de entrenamiento y validación. Durante el proceso de entrenamiento, el modelo ajustará sus pesos para aprender a reconocer patrones específicos asociados con objetos en las imágenes.

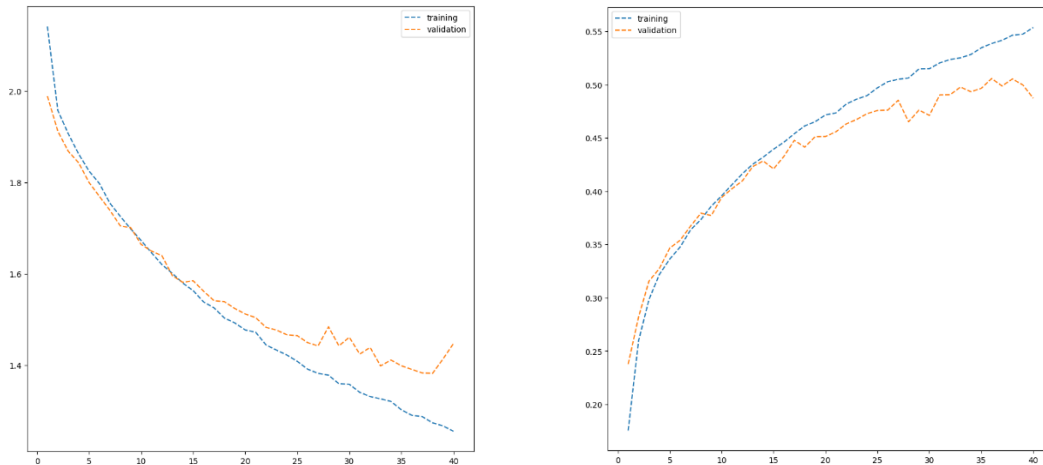
Evaluación del Modelo: Evaluaremos la efectividad del modelo mediante métricas de rendimiento como precisión, recall y F1-score en el conjunto de validación. Además, realizaremos validación cruzada y ajuste fino del modelo para asegurar su robustez y capacidad de generalización.

Análisis del procedimiento y resultados

Al realizar el análisis, limpieza y transformación de los datos observamos inicialmente que CIFAR-10 contiene imágenes de 10 clases diferentes, cada una representada por 5000 imágenes de entrenamiento y 1000 imágenes de prueba. Esto sugiere un conjunto de datos equilibrado que puede ser útil para entrenar un modelo de clasificación multiclase.

La estandarización de los datos (normalización entre 0 y 1) también es una buena práctica para el preprocesamiento de imágenes antes de alimentarlas a una red neuronal, ya que ayuda a mejorar la estabilidad del entrenamiento y a que el modelo converja más rápidamente. La arquitectura de la red neuronal creada inicialmente parece ser relativamente simple pero adecuada para comenzar, consta de varias capas densas con activación ReLU en las capas ocultas y activación softmax en la capa de salida, lo que sugiere que estás trabajando en una tarea de clasificación multiclase. El modelo tiene un total de 1,662,510 parámetros entrenables, lo cual indica que tiene cierta complejidad para aprender representaciones útiles de los datos.

Pasamos a entrenar la red neuronal observando gráficamente la pérdida durante el entrenamiento y validación de los datos:

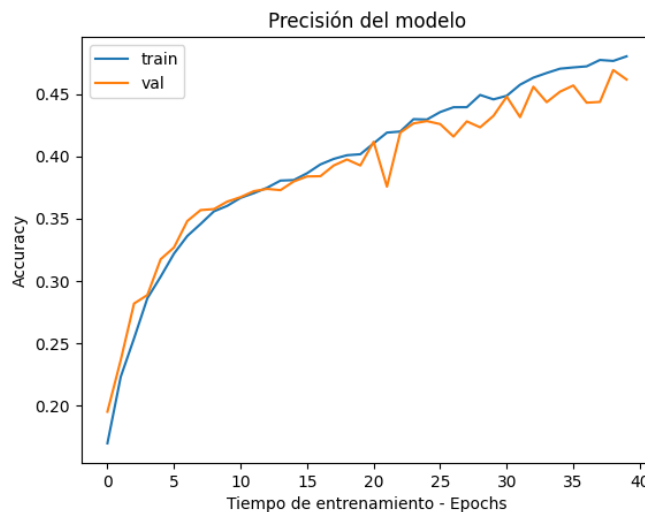


El entrenamiento de la red neuronal parece estar progresando a través de las épocas. Observamos que tanto la pérdida de entrenamiento como la pérdida de validación disminuyen inicialmente, lo que indica que el modelo está aprendiendo de los datos. Sin embargo, a medida que avanza el entrenamiento, vemos una tendencia a la estabilización o incluso a un ligero aumento en la pérdida de validación, mientras que la pérdida de entrenamiento sigue disminuyendo. Esto sugiere que el modelo puede estar sobreajustando los datos de entrenamiento y no generalizando tan bien con los datos de validación.

En la gráfica mostrada, podemos ver claramente la evolución de la pérdida tanto en el conjunto de entrenamiento como en el de validación a lo largo de las épocas, lo que nos brinda una visión general del rendimiento y la capacidad de generalización del modelo.

Se realizan predicción para compararlas con las etiquetas reales de las imágenes del conjunto de prueba. Esta comparación nos permite evaluar el rendimiento del modelo en términos de su capacidad para predecir correctamente las clases de las imágenes.

Precisión del modelo



Aquí se crea un gráfico que contrasta las curvas de precisión durante el proceso de entrenamiento y validación a lo largo del tiempo de entrenamiento - Epochs. La precisión es una métrica fundamental para evaluar el rendimiento de un modelo de clasificación. Analizar estas curvas facilita la identificación de mejoras en la capacidad predictiva del modelo, así como la detección de posibles problemas de sobreajuste o subajuste. En esta representación, se nota que la precisión se

mantiene relativamente constante entre el epochs 16 y 18, sugiriendo una estabilidad en el rendimiento del modelo en ese intervalo de entrenamiento.

Conclusiones

Modelo de Shallow Learning: 0.1

- En términos generales, el modelo con regularización tiene un rendimiento superior en términos de precisión en comparación con el modelo de Shallow Learning. Un mayor valor de precisión indica que el modelo con regularización ha clasificado correctamente una mayor proporción de ejemplos en comparación con el modelo de Shallow Learning.
- Dado que el modelo desarrollado identifica con precisión las imágenes cuando se ejecutan de manera aleatoria "Graficando Clase real vs. Clase predecidas", se puede concluir que ha tenido éxito en la detección de objetos en la mayoría de los casos.
- Las gráficas de pérdida durante el entrenamiento y la validación muestran una disminución constante, lo que indica que el modelo ha aprendido de manera efectiva y ha convergido hacia una solución. No se observan signos de sobreajuste.
- Al observar estas métricas, podemos concluir que el modelo con regularización tiene un desempeño aceptable, con una precisión y un recall variados según la clase. Sin embargo, aún hay margen de mejora, especialmente en las clases con puntuaciones más bajas en precisión y recall. Esto sugiere que ajustar los hiperparámetros del modelo o probar con diferentes técnicas de preprocesamiento podría mejorar su rendimiento.
- El éxito del modelo en la detección de objetos sugiere que la propuesta de solución, que incluye el uso de redes neuronales convolucionales y un enfoque meticuloso en el preprocesamiento y la arquitectura del modelo, ha sido efectiva. Por lo tanto, se recomienda la aplicación de este modelo para mejorar la capacidad del sistema de visión del vehículo y así identificar y clasificar objetos en tiempo real.

Recomendaciones

- Dado que el modelo con regularización ha mostrado un desempeño aceptable pero con variaciones en precisión y recall según la clase, se recomienda realizar una optimización exhaustiva de los hiperparámetros. Esto podría incluir ajustes en la tasa de aprendizaje, el tamaño del lote, el número de épocas de entrenamiento, entre otros, con el fin de mejorar el rendimiento en las clases con puntuaciones más bajas.
- Para mejorar aún más la capacidad de detección de objetos del modelo, se sugiere explorar técnicas avanzadas de preprocesamiento de imágenes como el uso de técnicas de aumento de datos, normalización adicional, o incluso la aplicación de filtros específicos para resaltar características relevantes en las imágenes que puedan estar siendo pasadas por alto en la detección actual.

Referencias

Documento en línea: <https://idus.us.es/bitstream/handle/11441/90004/Centeno%20Franco%20Alba%20TFG.pdf>