

## Parte 1

### a. ¿Cuál es la diferencia principal entre ambas arquitecturas?

Siendo ambas CNN tienen sus diferencias. En el caso de AlexNet tiene más capas de convolución y fully conectadas que LeNet. Luego LeNet usa la función de activación sigmoide mientras que AlexNet usa ReLU. Esto también hace que el ingreso o la entrada de los modelos varíe, por ejemplo en LeNet solamente se pueden ver imágenes en blanco y negro de 32 píxeles mientras que en AlexNet imágenes a color RGB de 224 píxeles.

### b. ¿Podría usarse LeNet-5 para un problema como el que resolvió usando AlexNet? ¿Y viceversa?

Si pero debe de primero considerar que la capacidad que se va a requerir va a ser distinta. También el size de los valores de entrada va a cambiar pues debemos de organizar y corregir los píxeles de las imágenes y asegurarnos que las mismas tengan color blanco y negro únicamente. Hecho esto en teoría si se puede usar LeNet para un problema resuelto con AlexNet.

En cuanto al caso de viceversa, si se puede pues AlexNet no cuenta con problema en cuanto a los colores pero si debemos de ver el size de los píxeles de entrada y considerar que el modelo va a ser mucho más pesado computacionalmente por las capas extra que tiene a LeNet.

### c. Indique claramente qué le pareció más interesante de cada arquitectura

Principalmente la complejidad de una red puede aumentar siempre tomando los avances realizados con otra como en este caso se hicieron avances en AlexNet tomando en cuenta la arquitectura pionera de LeNet. También factores como lo es que ambas tienen una gran diferencia de capas y funciones de activación lo que hace que el modelo termine requiriendo de muchos más recursos. Ambos de estos modelos fueron importantes para avances en deep learning pero también es de notar que gracias a la transferencia de aprendizaje estos han ayudado a popularizar el uso de estos modelos lo que lleva a avances como lo que hoy en día usa GPT como base para lograr modelos complejos y mucho más robustos con mayor capacidad.

## Parte 2

### 1. Investigue e indique en qué casos son útiles las siguientes arquitecturas, agregue imágenes si esto le

ayuda a una mejor comprensión

a. GoogleNet (Inception):

La arquitectura GoogleNet, también conocida como Inception, es útil en casos en los que se requiere un alto rendimiento en tareas de clasificación de imágenes con una cantidad limitada de parámetros. Esta arquitectura se caracteriza por su módulo Inception, que utiliza múltiples tamaños de filtros de convolución en paralelo para capturar características de diferentes escalas. Es especialmente útil en la clasificación de imágenes donde se deben identificar objetos en diferentes escalas y posiciones. La arquitectura Inception V3, por ejemplo, es conocida por su eficiencia y precisión en la clasificación de imágenes.

b. DenseNet (Densely Connected Convolutional Networks):

DenseNet es útil en casos donde se necesita aprovechar al máximo la información de las capas previas en una red convolucional. En lugar de sumar o concatenar solo las salidas de capas anteriores, DenseNet conecta cada capa con todas las capas posteriores. Esto fomenta un flujo de información denso y facilita el entrenamiento de redes profundas sin el problema de desvanecimiento del gradiente. DenseNet es particularmente eficaz en conjuntos de datos pequeños y en tareas de segmentación de imágenes.

c. MobileNet:

MobileNet es útil en aplicaciones donde se requiere una red neuronal ligera y eficiente en términos de consumo de recursos, como dispositivos móviles y sistemas embebidos. Esta arquitectura utiliza operaciones de convolución separable para reducir drásticamente el número de parámetros y la cantidad de cálculos necesarios, mientras mantiene un buen rendimiento en tareas de clasificación de imágenes. Es ideal para aplicaciones de visión por computadora en dispositivos con recursos limitados.

d. EfficientNet:

EfficientNet es útil cuando se busca una arquitectura que sea eficiente en términos de consumo de recursos computacionales y al mismo tiempo sea altamente efectiva en tareas de clasificación de imágenes. Esta arquitectura utiliza un enfoque de escalado compuesto que equilibra el ancho, la profundidad y la resolución de la red para adaptarse a diferentes tamaños de datos y requisitos de rendimiento. Es especialmente útil en aplicaciones donde se necesitan redes altamente escalables y precisas.

## 2. ¿Cómo la arquitectura de transformers puede ser usada para image recognition?

Si bien las arquitecturas de Transformers se diseñaron originalmente para el procesamiento de lenguaje natural, se han adaptado con éxito para tareas de reconocimiento de imágenes mediante el uso de una técnica llamada "Transformers Visuales" o "Vision Transformers" (ViT). Esto sirve al tomar una foto y generar matrices donde se procesa la información de la

forma que se procesa el texto. Esto tambien permite hacer embedding de posicion de las imagenes capturando la inforamcion en parches de las imagenes traducidas al modelo. Esto nos permite luego poder hacer inferencias en nuevas imagenes.