Ejercicios de Aprendizaje Profundo -Parte 1

Inteligencia Artificial

1 Lista de ejercicios

1. Responde las siguientes preguntas:

   (a) Una red neuronal de múltiples capas con funciones de activación lineal es equivalente a un perceptrón de una sola capa que usa la misma función de error en la capa de salida y tiene el mismo número de entradas.

   (b) Las redes neuronales convolucionales no varían en rotación.

   (c) César ahora ha cambiado a redes neuronales de múltiples capas y nota que el error de entrenamiento está disminuyendo y converge a un mínimo local. Luego, cuando prueba los nuevos datos, el error de prueba es anormalmente alto. ¿Qué es lo que probablemente está saliendo mal y qué le recomiendas que haga?. Indica 3 opciones.

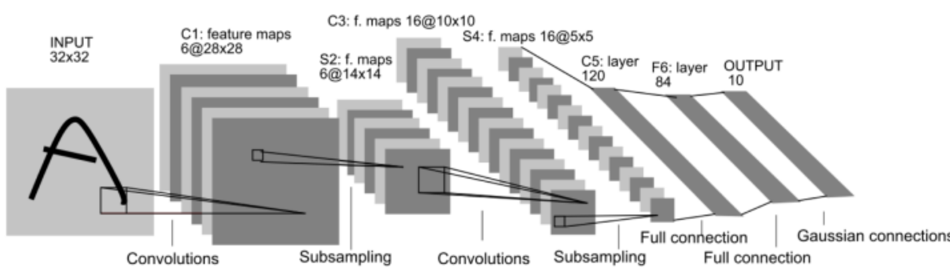
   (d) Una red neuronal con múltiples capas ocultas y nodos sigmoides puede formar límites de decisión no lineales.

   (e) Todas las redes neuronales calculan funciones no convexas de sus parámetros.

   (f) Las redes neuronales convolucionales generalmente tienen menos parámetros libres en comparación con las redes neuronales completamente conectadas.

   (g) Supongamos que cuando estás entrenando una red neuronal convolucional, encuentras que la pérdida de entrenamiento simplemente no disminuye después de la inicialización. ¿Qué podrías intentar para solucionar este problema?. Menciona 3 posibles soluciones.

2. Aquí está la arquitectura histórica de la red neuronal convolucional LeNet de Yann LeCun.para la clasificación de dígitos. Aquí, la capa `INPUT` toma una imagen de 32×32 y la capa `OUTPUT` produce 10 salidas. La notación $6@28 \times 28$ significa 6 matrices de tamaño 28×28.



   Si los parámetros de una capa determinada son los pesos que se conectan a sus entradas,

   (a) Dado que el tamaño de entrada es 32×32 y el tamaño de la Capa 1 es 28×28, ¿cuál es el tamaño del filtro convolucional en la primera capa (es decir, a cuántas entradas está conectada cada neurona)?.

   (b) ¿Cuántos parámetros independientes (peso y sesgo) hay en la capa C1?

(c) ¿Cuántos parámetros independientes (peso y sesgo) hay en la capa C3?

(d) ¿Cuántos parámetros independientes (ponderación y sesgo) hay en la capa F6?

3. Para mostrar que las redes de retropropagación pueden dividir conjuntos separables no linealmente, entrena la función XOR. Los datos de entrenamiento para esto son: $((0,0),0),((0,1),1),((1,0),1),((1,1),0)$.

   (a) Crea una red con dos neuronas cada una en la capa de entrada y la capa oculta, y una neurona en la capa de salida, y entrene esta red.

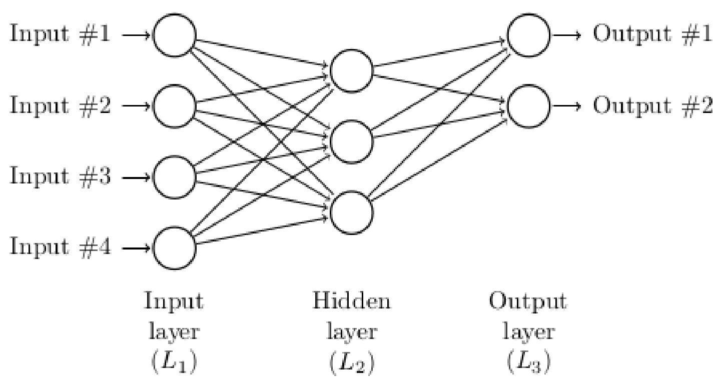
   (b) Ahora elimine las neuronas ocultas y conecta la capa de entrada directamente a las neuronas de salida. ¿Qué observas?.

4. Responde las siguientes preguntas:

   (a) Muestra que cualquier red de retropropagación multicapa con una función de activación lineal es igualmente poderosa que una de dos capas. Para ello basta con mostrar que las sucesivas ejecuciones de mapeos lineales es un mapeo lineal.

   (b) Muestra que una red de retropropagación de dos capas con cualquier función de activación estrictamente monótona no es más poderosa para tareas de clasificación que una sin una función de activación o con una función de activación lineal.

5. Considera una red de tres capas completamente conectada con n_1, n_2, n_3 neuronas en tres capas respectivamente. Las entradas se introducen en la primera capa. La pérdida es el error cuadrático medio E y la no linealidad es una función sigmoidea. Sea t el vector de etiqueta de tamaño n_3. Sea cada vector de salida de capa y_i y el vector de entrada sea z_i, ambos de tamaño n_i. Sea W_{ii+1} el peso entre la capa i y la capa $i+1$. El j-ésimo elemento en y_i está definido por y_i^j, lo mismo para z_i^j. El peso que conecta la neurona k-ésima y l-ésima en las capas i, $i+1$ está definido por W_{ii+1}^{kl}.



Resumen de la notación

- σ denota la función de activación para L_2 y L_3, $\sigma(x) = \frac{1}{1+e^{-x}}$. No se aplica ninguna activación a la capa de entrada.
- $z_i^{(j)} = \sum_{k=1}^{P} W_{i-1i}^{kj} x_{i-1}^{(k)}$
- $y_i^{(j)} = \sigma(\sum_{k=1}^{P} W_{i-1i}^{kj} x_{i-1}^{(k)})$

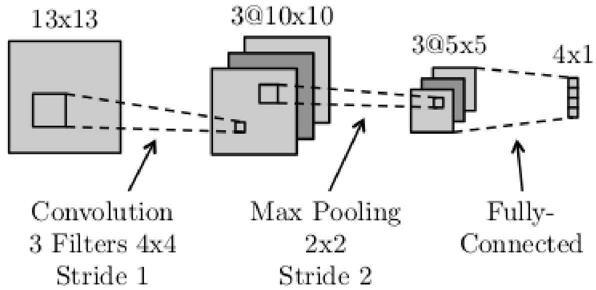
(a) Encuentra $\frac{\partial E}{\partial z_3^j}$.

(b) Encuentra $\frac{\partial E}{\partial y_2^k}$ en términos de elementos en W_{23} y $\frac{\partial E}{\partial z_3^j}$.

(c) Encuentra $\frac{\partial E}{\partial w_{23}^{kj}}$ en términos de y_2^k, y_3^j y t^j.

(d) Si la entrada a una neurona en la capa max-pooling es x y la salida es $y = \max(x)$, calcula $\frac{\partial y}{\partial x_i}$.

6. A continuación se muestra un diagrama de una pequeña red neuronal convolucional que convierte una imagen de 13×13 en 4 valores de salida. La red tiene las siguientes capas-operaciones desde la entrada hasta la salida: convolución con 3 filtros, max pooling, ReLu y finalmente, una capa completamente conectada. Para esta red no utilizaremos ningún parámetro de bias-offset (b). Responda las siguientes preguntas sobre esta red.



(a) ¿Cuántos pesos en la capa convolucional necesitamos aprender?

(b) ¿Cuántas operaciones ReLu se realizan en el paso forward?

(c) ¿Cuántos pesos debemos aprender para toda la red?

(d) Explica el siguiente resultado: una red neuronal completamente conectada con capas del mismo tamaño que la red anterior ($13 \times 13 \rightarrow 3 \times 10 \times 10 \rightarrow 3 \times 5 \times 5 \rightarrow 4 \times 1$) puede representar cualquier clasificador que pueda representar la red convolucional anterior.

(e) ¿Cuál es la desventaja de una red neuronal completamente conectada en comparación con una red neuronal convolucional con capas del mismo tamaño?.