06 - ¿Por qué usamos el gradiente para entrenar Redes Neuronales? Ejemplo - Deep Learning - (Parte 6) - Transcript

En el vídeo anterior presentamos el algoritmo "Grate In Decent" el cual es utilizado para ajustar los valores de los parámetros de nuestra red neuronal y así disminuir el valor de la función de costo.

En este vídeo vamos a presentar un ejemplo muy sencillo que nos permitirá crear una intuición de por qué al momento de cambiar los parámetros W o B por un valor pequeño H, el valor a la salida de la función es afectado.

Entonces empezamos y vamos a resolver este ejemplo muy sencillo que nos va a permitir de forma muy intuitiva y simple darnos cuenta cómo este valor de la salida como este valor acá va a ser afectado por cambios muy pequeños de estos valores.

Después de hacer este proceso en el siguiente vídeo vamos a aplicar estos conceptos pero ya para una red neuronal muy muy pequeñita pero que nos permita ver cómo la función de pérdida es afectada mediante el ajuste de los dadores W y B.

Suponemos entonces que este es un equivalente a nuestra neuronal artificial que realiza la multiplicación Wx y le suma el valles simplemente hemos separado en nodos en esta gráfica computacional.

Entonces tenemos que aquí vamos a llamar este valor al punto a la salida de la multiplicación vamos a llamarle t y aquí tendríamos que t va a ser igual a -10 ya que es la multiplicación de W por x igual a 5 y a la salida aquí tendríamos que es -10 y le estamos sumando un valor de b igual a 3 entonces tendríamos que nuestra salida vamos a llamarle a esta salida z es igual a -7.

Lo que queremos encontrar es que tanto cambia z con respecto a W con respecto a x y con respecto a b es decir si variamos W por ejemplo un poquito que tanto afecta la salida z en este caso es una neurona muy sencillita de solamente se realiza una operación lineal pero este va a ser el mismo principio para la función de pérdida de una red neuronal mucho más compleja entonces para esto lo que vamos a hacer es uso de la regla de la cadena que es un concepto fundamental de cálculo diferencial lo que vamos a ir haciendo es ir calculando la derivada del número que tenemos del último número con respecto a las entradas de este módulo y así aplicando la regla de la cadena vamos a ir haciendo una retro propagación del gradiente por lo tanto este algoritmo se conoce como back propagation vamos a hacer el ejemplo y esto nos va a permitir entender de forma mucho más sencilla esto que estoy explicando y vamos a empezar por definir algunas derivadas muy sencillas de esta arquitectura por ejemplo tenemos t que es la salida de este módulo de multiplicación t tenemos que está dado por la multiplicación de W por x tenemos que la derivada de t con respecto a W es igual a x y la derivada de t con respecto a x es igual a W por lo tanto en este caso en particular la derivada de t con respecto a W será igual a x y x tiene un valor de 5 y la derivada de t con respecto a x sería igual a W y W tiene un valor de -2 por lo tanto aquí sería -2 de igual forma si pensamos que aquí aquí aquí la entrada de este módulo a este módulo de suma es t y aquí es la entrada la entrada es b definimos que z está dado por t más b que tanto cambia z si t cambia un cierto valor iré igual forma que tanto cambia z si b cambia un cierto valor que sería la derivada de z con respecto a b entonces vamos a escribir esto y tendríamos que la derivada de z con respecto a t sería igual a 1 podemos encontrar tablas con las reglas de derivadas pero bueno la derivada de una suma en este caso sería la derivada esta es una constante que no depende de t por lo tanto sería 0 y esta sería la derivada de t con respecto a t lo cual es 1 y entonces tenemos finalmente que la derivada de z con respecto a b también sería 1 por los mismos motivos que el caso de la derivada de z con respecto a t y ahora lo que vamos a hacer entonces es obtener estos valores de la derivada de z con respecto a b la derivada de z con respecto a x y la derivada de z con respecto a b y para esto vamos a utilizar estas derivadas que acabamos de definir aquí la derivada de z con respecto a w estaría dada por la derivada de z con respecto a t por la derivada de t con respecto a w y si sustituimos los valores que tenemos aquí tenemos que la derivada de t con respecto a perdón la derivada de z con respecto a t es igual a 1 y la derivada de t con respecto a w es igual a 5 por lo tanto la derivada de w de z con respecto a w es igual a 5 de forma similar la derivada de z con respecto a x la calculamos como la derivada de z con respecto a t por la derivada de t con respecto a x y esto de nueva cuenta es 1 esta es la derivada de z con respecto a t por la derivada de t con respecto a x que es -2 y esto nos da un valor de -2 finalmente la derivada de z con respecto a b está dada por la derivada de z con respecto a b la cual ya teníamos calculada acá que es igual a 1 entonces si analizamos cada uno de estos módulos tenemos un gradiente en este punto que es el resultado de la derivada de la salida en este caso z con respecto a dt a la entrada tenemos aquí también un gradiente que sería la derivada de z con respecto a db de igual forma aquí en este punto tenemos un gradiente que es la derivada de t con respecto a dw y aquí un gradiente que es la derivada de t con respecto a dx y en este caso a la salida es importante notar que tendríamos posiblemente un gradiente si éste estuviera conectado a otros puntos tendríamos un gradiente que llega de regreso en este caso como no está conectado a ningún otro lado la derivada en este punto no es otra cosa que la derivada de z con respecto a sí mismo y esto es igual a 1 sería la derivada aquí a la salida para obtener estos valores de dz de w una forma de visualizar esto es que tenemos un gradiente que llega aquí que llega desde arriba que llega de las neuronas que están posteriores y ese y entonces lo que queremos hacer es calcular la derivada de este dz con respecto a b aquí fluye un gradiente de regreso estrictamente este gradiente sería la derivada de z con respecto a dz por la derivada de z con respecto a db ese sería el gradiente que está yendo aquí de regreso en este sentido y ese gradiente habíamos dicho que en este caso es de uno de igual forma aquí en esta suma en esta poración de suma fluye un gradiente de regreso y ese gradiente de regreso es el gradiente de dz con respecto a dt el cual habíamos dicho de nuevo que es un valor de uno estrictamente también este valor de dz con respecto a dt sería igual a este gradiente de acá de z con respecto a de z por el gradiente de z con respecto a d t no ponemos esto porque esto es igual a uno no ponemos de z con respecto a de z porque esto es igual a uno pero estrictamente lo deberíamos de estar escribiendo igual aquí de z con respecto de z y de z con respecto a db y así fluye un gradiente de regreso que es de uno pero ahora lo que queremos es encontrar el gradiente que va de regreso para w y para x y entonces lo que tenemos es que por ejemplo el gradiente aquí que va de regreso sería este gradiente de z con respecto a dw que mencionamos que está dado como el resultado de la regla de la cadena que es de z con respecto a de t que es este gradiente que va aquí de regreso que ya dijimos que es uno por el gradiente local y el gradiente local dijimos que era este de t con respecto a de w tal forma que el gradiente que regresa aquí es 5 aquí el gradiente que regresa a la x también lo calculamos de forma similar que es el gradiente local de dt con respecto a de x es esta parte multiplicado por el gradiente que está llegando a esta unidad que es de z con respecto a de t y así se calcula que este gradiente es de -2 y aquí fluye un gradiente de -2 y esos son los gradientes que van a regresar a estos puntos ahora de qué nos sirve esto tomando este valor de h igual a 1 vamos a decir que afectamos w por ese valor de 1 w va a ser igual a hora w más h por lo tanto w este nuevo valor de w va a ser igual a -1 entonces vamos a hacer los cálculos nuevamente todos estos valores pero ahora asumiendo que w ya no vale -2 sino vale -1 porque lo estamos incrementando por este valor de 1 y es un valor que seleccione arbitrariamente simplemente para hacer un ejemplo muy muy sencillo de ahora sería igual a el valor de w que ahora es -1 por el valor de x que sigue siendo 5 y esto nos va a dar -5 y este es el valor que entra a esta suma de aquí tenemos que z estaría dado por la suma de t + b z de t + b como como definimos aquí en esta parte entonces este nuevo valor de z sería igual a t que ahora es -5 más b que ahora vale bueno sigue valiendo 3 entonces este valor nos va a dar -5 más 3 igual a -2 y nuestro nuevo valor de z es -2 ahora vamos a analizar este resultado originalmente nuestro valor de z era -7 para un valor de w igual a -2 pero la derivada de z con respecto a w nos dice que es 5 y este valor 5 implica que si incrementamos w por un valor h que en este caso h fue de 1 nuestro valor de z se va a incrementar por un valor de cinco veces ese valor h recordemos la derivada de z con respecto a w nos dice que tanto cambia z si variamos el valor w y vamos a verlo lo que hicimos nosotros fue incrementar w por un valor de 1 asumimos h igual a 1 entonces w se incrementó una unidad y esperamos que z se incremente cinco veces una unidad es decir 5 si pensamos -7 más 5 5 por 1 bueno el valor de h esto nos da igual a -2 y repitiendo entonces el flujo de la información de los datos en este modelo de gráfica computacional tenemos que con un nuevo valor de w igual a -1 si repetimos t ahora vale -5 y luego calculamos z efectivamente nos da -2 lo que nos demuestra que como esperábamos z se incrementó una cantidad igual a cinco veces h vamos a hacer un ejemplo ahora con x y vamos a repetir el experimento de nuevo cuenta con h igual a 1 entonces x ahora sería igual a x más h y esto sería igual a 5 más 1 sería igual a 6 nuestro nuevo valor de x va a ser 6 y repetimos t va a ser entonces igual a w sigue teniendo este valor original de -2 a w no lo modificamos pero ahora x es un valor de 6 esta multiplicación nos da -12 y de aquí calculamos nuestro nuevo valor de z el cual está dado de nueva cuenta como t más b y z ahora estaría dado por t que es -12 y más b no lo hemos cambiado sigue siendo 3 entonces nuestra nueva z está dada por un valor de -9 este sería nuestro valor de z vamos a ver si este valor coincide con la expectativa dada la derivada de z con respecto a x según la derivada de z con respecto a x nos da que es -2 por lo que esperaríamos que si variamos x por un valor igual a h z debería ser incrementada por un valor igual a -2 h y aquí incrementamos x por un valor de 1 por lo tanto esperaríamos que z en este caso como la derivada es -2 es negativa esperaríamos que z se incremente por un valor de -2 h o sea es decir que se decremente por una cantidad de 2 h validando nuestra operación lo que observamos es que efectivamente z tenía un valor de -7 si este valor -7 más -2 por h que en este caso h es 1 esto nos da un valor igual a -9 el cual podemos ver que coincide con el resultado que tenemos aquí concluimos que el resultado de la derivada de z con respecto a x que es -2 realmente coincide con la evaluación de la gráfica computacional cuando incrementamos x por un valor de 1 en este caso y así entonces vemos que hay una relación entre qué tanto cambia nuestra función en este caso f z pero lo vamos a hacer para nuestra función de costo y ver qué tanto cambia nuestra función de costo con respecto a nuestras a nuestros parámetros w y nuestros parámetros b en este vídeo hemos presentado un ejemplo muy sencillo que nos permite crear cierta intuición de por qué al ajustar los valores de nuestros parámetros w y b el valor a la salida de la función también es afectado y lo que vamos a hacer en el siguiente vídeo es crear un ejemplo un poco más complejo aún siendo muy muy sencillo para poder realizarlo de forma didáctica y lo que nos va a permitir este ejemplo es ver cómo efectivamente al cambiar nuestros parámetros w y b siguiendo el algoritmo de gradient descent el valor de la función de costo va a ir disminuyendo y una vez hecho esto vamos a tener todas las herramientas necesarias para poder construir un modelo de red neuronal que sea más complejo de más de una capa y con esto ya vamos a poder continuar finalmente a la implementación de nuestro primer modelo de red neuronal en código entonces continuamos [AUDIO\_EN\_BLANCO]