 

**Universidad Autónoma de Sinaloa**

**Ingeniería de Software**

**4-03**

**Redes Neuronales**

**C8-T4: Pruning Neural Network**

**Elaborado por:**

**Valdez Villegas Evie**

**Profesor: M.C. Edgar Omar Pérez Contreras**

**26 de Mayo del 2024**

**Índice**

[Introduccion 3](#_Toc167641087)

[Entendiendo la Poda 4](#_Toc167641088)

[Poda de Conexiones 4](#_Toc167641089)

[Poda de Neuronas 4](#_Toc167641090)

[Mejorando o Degradando el Rendimiento 4](#_Toc167641091)

[Algoritmos de Poda 5](#_Toc167641092)

[Poda Incremental 5](#_Toc167641093)

[Poda Selectiva 6](#_Toc167641094)

[Conclusión 7](#_Toc167641095)

## Introduccion

En los capítulos 6 y 7, vimos que se pueden utilizar entrenamiento simulado y algoritmos genéticos para entrenar redes neuronales. Estas dos técnicas emplean varios algoritmos para ajustar mejor los pesos de una red neuronal al problema al que se aplica. Sin embargo, estas técnicas no hacen nada para ajustar la estructura de la red neuronal. En este capítulo, examinaremos dos algoritmos que pueden usarse para modificar realmente la estructura de una red neuronal. Esta modificación estructural generalmente no mejorará la tasa de error de la red neuronal, pero puede hacer que la red neuronal sea más eficiente. La modificación se logra analizando cuánto contribuye cada neurona a la salida de la red neuronal. Si la conexión de una neurona en particular a otra neurona no afecta significativamente la salida de la red neuronal, la conexión se eliminará. A través de este proceso, se eliminan las conexiones y las neuronas que tienen un impacto marginal en la salida.

Este proceso se llama poda. En este capítulo, examinaremos cómo se realiza la poda. Comenzaremos examinando el proceso de poda en mayor detalle y discutiremos algunos de los métodos populares. Finalmente, este capítulo concluirá proporcionando dos ejemplos que demuestran la poda.

## Entendiendo la Poda

La poda es un proceso utilizado para hacer que las redes neuronales sean más eficientes. A diferencia de los algoritmos genéticos o el entrenamiento simulado, la poda no aumenta la efectividad de una red neuronal. El objetivo principal de la poda es disminuir la cantidad de procesamiento requerido para usar la red neuronal.

La poda puede ser especialmente efectiva cuando se realiza en una red neuronal grande que tarda demasiado tiempo en ejecutarse. La poda funciona analizando las conexiones de la red neuronal. El algoritmo de poda busca conexiones individuales y neuronas que se pueden eliminar de la red neuronal para que funcione de manera más eficiente. Al podar conexiones innecesarias, la red neuronal puede hacerse más rápida. Esto permite que la red neuronal realice más trabajo en una cantidad de tiempo determinada. En las siguientes dos secciones examinaremos cómo podar tanto conexiones como neuronas.

## Poda de Conexiones

La poda de conexiones es central para la mayoría de los algoritmos de poda. Las conexiones individuales entre las neuronas se analizan para determinar qué conexiones tienen el menor impacto en la efectividad de la red neuronal. Uno de los métodos que examinaremos eliminará todas las conexiones que tengan un peso por debajo de un valor umbral específico. El segundo método evalúa la efectividad de la red neuronal a medida que se consideran ciertos pesos para su eliminación. Las conexiones no son lo único que se puede podar. Al analizar qué conexiones se podaron, también podemos podar neuronas individuales.

## Poda de Neuronas

La poda se enfoca principalmente en las conexiones entre las neuronas individuales de la red neuronal. Sin embargo, las neuronas individuales también se pueden podar. Uno de los algoritmos de poda que examinaremos más adelante en este capítulo está diseñado para podar tanto neuronas como conexiones.

Para podar neuronas individuales, se deben examinar las conexiones entre cada neurona y las otras neuronas. Si una neurona en particular está rodeada completamente por conexiones débiles, no hay razón para mantener esa neurona. Si aplicamos los criterios discutidos en la sección anterior, podemos terminar con neuronas que no tienen conexiones. Esto se debe a que todas las conexiones de la neurona fueron podadas. Tal neurona puede ser podada a su vez.

## Mejorando o Degradando el Rendimiento

Es posible que podar una red neuronal mejore su rendimiento. Cualquier modificación en la matriz de pesos de una red neuronal siempre tendrá algún impacto en la precisión de los reconocimientos realizados por la red neuronal. Una conexión que tiene poco o ningún impacto en la red neuronal puede estar degradando la precisión con la que la red neuronal reconoce patrones. Eliminar una conexión débil puede mejorar la salida general de la red neuronal.

Desafortunadamente, también es posible disminuir la efectividad de la red neuronal mediante la poda. Por lo tanto, siempre es importante analizar la efectividad de la red neuronal antes y después de la poda. Dado que la eficiencia es el principal beneficio de la poda, debe tener cuidado de evaluar si una mejora en el tiempo de procesamiento vale una disminución en la efectividad de la red neuronal. El programa en el ejemplo que examinaremos más adelante en este capítulo evaluará la efectividad general de la red neuronal tanto antes como después de la poda. Esto nos dará una idea de qué efecto tuvo el proceso de poda en la efectividad de la red neuronal.

## Algoritmos de Poda

Ahora revisaremos exactamente cómo se lleva a cabo la poda. En esta sección examinaremos dos métodos diferentes para la poda. Estos dos métodos funcionan de manera algo opuesta. El primer método, la poda incremental, funciona aumentando gradualmente el número de neuronas ocultas hasta que se haya obtenido una tasa de error aceptable. El segundo método, la poda selectiva, funciona tomando una red neuronal existente y disminuyendo el número de neuronas ocultas siempre que la tasa de error siga siendo aceptable.

## Poda Incremental

La poda incremental es un enfoque de prueba y error para encontrar un número apropiado de neuronas ocultas.

El algoritmo de poda incremental comienza con una red neuronal no entrenada. Luego intenta entrenar la red neuronal muchas veces. Cada vez, usa un conjunto diferente de neuronas ocultas.

El algoritmo de entrenamiento incremental debe proporcionarse con una tasa de error aceptable. Busca la red neuronal con el menor número de neuronas ocultas que haga que la tasa de error caiga por debajo del nivel deseado. Una vez que se encuentra una red neuronal que se pueda entrenar para que caiga por debajo de esta tasa, el algoritmo está completo.

Como viste en el capítulo 5, "Redes Neuronales de Retropropagación de Alimentación", a menudo es necesario entrenar durante muchos ciclos antes de encontrar una solución. El algoritmo de poda incremental requiere que se complete toda la sesión de entrenamiento muchas veces. Cada vez que se agrega una nueva neurona a la capa oculta, la red neuronal debe volver a entrenarse. Como resultado, puede llevar mucho tiempo ejecutar el algoritmo de poda incremental.

La red neuronal se entrenará para diferentes números de neuronas ocultas, comenzando inicialmente con una sola neurona. Debido a que la tasa de error no disminuye lo suficientemente rápido, la red neuronal de una sola neurona se abandonará rápidamente. Se pueden usar cualquier número de métodos para determinar cuándo abandonar una red neuronal. El método que se usará en este capítulo es verificar la tasa de error actual después de intervalos de 1,000 ciclos. Si el error no disminuye en un punto porcentual, entonces se abandonará la búsqueda. Esto nos permite abandonar rápidamente tamaños de capa oculta que son demasiado pequeños para la tarea prevista.

Una ventaja del algoritmo de poda incremental es que generalmente creará redes neuronales con menos neuronas ocultas que los otros métodos. La mayor desventaja es la cantidad de tiempo de procesador que lleva ejecutar este algoritmo. Ahora que te hemos presentado el algoritmo de poda incremental, examinaremos el algoritmo de poda selectiva.

## Poda Selectiva

El algoritmo de poda selectiva difiere del algoritmo de poda incremental en varios aspectos importantes. Una de las diferencias más notables es el estado inicial de la red neuronal. No se requirió ningún entrenamiento antes de comenzar el algoritmo de poda incremental. Este no es el caso con el algoritmo de poda selectiva. El algoritmo de poda selectiva funciona examinando las matrices de pesos de una red neuronal previamente entrenada. Luego, el algoritmo de entrenamiento selectivo intentará eliminar neuronas sin interrumpir la salida de la red neuronal.

## Conclusión

¡Y eso concluye el resumen del capítulo sobre la poda de redes neuronales! Este proceso de poda ofrece una forma de mejorar la eficiencia de las redes neuronales al eliminar conexiones y neuronas que tienen un impacto mínimo en la salida de la red. Con métodos como la poda incremental y selectiva, se puede ajustar la estructura de la red neuronal para hacerla más eficiente sin sacrificar significativamente su efectividad.

En resumen, la poda de redes neuronales nos muestra una lección valiosa: a veces, menos es más. Al eliminar conexiones y neuronas que contribuyen mínimamente al rendimiento de la red, podemos mejorar su eficiencia y velocidad de procesamiento sin comprometer su capacidad para realizar tareas complejas. Sin embargo, también debemos recordar que cada modificación tiene un impacto, y es crucial evaluar cuidadosamente si los beneficios de la poda superan cualquier pérdida en la precisión o efectividad de la red. Al igual que en la vida, encontrar el equilibrio adecuado entre eficiencia y rendimiento es fundamental para el éxito a largo plazo.