Algoritmos con tasas de aprendizaje adaptativo

Los algoritmos con tasas de aprendizaje adaptativo en redes neuronales son métodos que ajustan automáticamente la velocidad de aprendizaje durante el proceso de entrenamiento. La tasa de aprendizaje es un hiperparámetro crucial que determina el tamaño de los pasos que se dan para actualizar los pesos de la red durante el descenso del gradiente.

La adaptación de la tasa de aprendizaje busca mejorar la convergencia del modelo y evitar problemas como convergencia lenta o estancamiento.

Algunos algoritmos populares de tasas de aprendizaje adaptativo incluyen:

Adagrad (Adaptive Gradient Algorithm)

Ajusta la tasa de aprendizaje de cada parámetro según su historial de gradientes.

Los parámetros que han tenido gradientes grandes tendrán una tasa de aprendizaje reducida y viceversa. Es útil para problemas donde las características varían significativamente en escala.

RMSprop (Root Mean Square Propagation)

Mejora Adagrad al utilizar una media ponderada de los gradientes cuadrados recientes.

Mitiga el problema de que la tasa de aprendizaje disminuya demasiado rápidamente.

Adam (Adaptive Moment Estimation)

Combina elementos de Adagrad y RMSprop.
Mantiene una media móvil de primer y segundo orden de los gradientes.

Adapta la tasa de aprendizaje para cada parámetro y también introduce términos de momento.

Estos algoritmos, abordan la dificultad de elegir una tasa de aprendizaje fija que funcione bien para todas las partes del espacio de parámetros. La adaptación permite una convergencia más eficiente y, en muchos casos, puede mejorar la velocidad y la estabilidad del entrenamiento. La elección del algoritmo depende de la naturaleza del problema y puede requerir ajustes específicos según la arquitectura de la red y la tarea de aprendizaje.