

Problemas en el ajuste de la red

February 17, 2025

1 La red no llega a aprender. Lo hace mal en el conjunto de ajuste

Pueden pasar dos cosas:

1. la red es poco potente
2. el ajuste no avanza

Si por alguna razón has optado por una red pequeña, posiblemente estés en el primer caso. Si tu red es grande, puedes pensar que estás en el segundo.

En el primer caso sí se ve progreso de ajuste, pero es pequeño. En el segundo es que apenas se ve progreso, o se detiene muy pronto.

1.1 La red es poco potente

1. A lo mejor no puede por un problema de los datos. ¿Los has mirado? ¿Necesitas filtrar ruido? Comprueba el preproceso.
2. Auméntala. La potencia aumenta más con el número de capas; es la primera opción si estás muy lejos. Con el número de procesadores por capa también aumenta, así que si ya ves avances, puedes probar eso solo.
3. Utiliza arquitecturas más complejas. Según el tipo de problema, puedes considerar convolutivas, realimentadas, con atención, con conexiones que salten capas, con varias ramas,
4. Prueba distintas no linealidades: ¿la salida está acotada o no? ¿convendría que hubiera derivada no nula en una zona mayor?

1.2 El ajuste no avanza

1. Puede ser simplemente cuestión de paciencia. Si la red es grande y compleja, el ajuste va a ser lento. Insiste.
2. El preproceso también puede influir. Si la distribución no es gaussiana, ni tiene mucha cola, acota los rangos. Si es más o menos gaussiana o con mucha cola, acota las varianzas.

3. Puede que sea demasiado grande. Intenta una menor. Si la necesitas grande, aumenta poco a poco, usando las menores previas como punto de partida de pesos.
4. La inicialización puede ser defectuosa. Cámbiala. Echa un vistazo a los pesos cuando se bloquea. Si son grandes, inicializa más bajo. Si son pequeños, inicializa más alto.
5. Cambia de algoritmo
6. Cambia la constante de ajuste. Considera la posibilidad de que sea variable, típicamente o empezando un poco alta y descendiendo, o empezando baja, subiendo y luego bajando (Pytorch tiene un CosineLR para automatizar esto).
7. Si estás usando regularización (“weight decay”) bájalo
8. Si el gradiente lo aplicas con toda la muestra o con una parte elevada, prueba a hacerlo por lotes menores (“batch size”)
9. Prueba capas normalizadoras
10. Prueba conexiones que salten varias capas (“residual”)
11. No es lo más frecuente, pero asegúrate de que la función de error es adecuada y numéricamente aceptable. Puedes utilizar métricas de error que se apliquen a capas intermedias, si tu diseño lo permite.
12. Comprueba que no estás introduciendo variables inútiles, que todo lo que pueden dar es palos en las ruedas
13. Si tienes acceso a una red similar, usa parte de ella congelando esos pesos

1.3 El ajuste diverge

1. Baja la constante de ajuste
2. Aumenta la regularización (“weight decay”)
3. Si estabas usando lotes de ajuste para el ajuste (“batch size”) auméntalo
4. Cambia a algoritmos que limiten los gradientes
5. Inicializa más bajo

2 Hace bien el conjunto de ajuste, pero sobrea-justa

Empieza asegurándote bien. Comprueba que tienes una buena estimación del error en generalización.

1. ¿Puedes aumentar la muestra, aunque sea artificialmente? Quizá puedas no perder parte para medir generalización si utilizas una estimación astuta.
2. Puede ser demasiado grande. Redúcela: quita capas, procesadores, de cara o anulando sobre el diseño que tienes.
3. Usa regularización (“weight decay”)
4. Anula procesadores aleatoriamente (“dropout”)
5. Parametriza los pesos (que sigan una distribución, que sean compartidos, filtros separables, ...)
6. ¿Puede que haya demasiadas variables? Selecciona