UNAM

Universidad Nacional Autónoma de México

INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS

Licenciatura en Ciencia de Datos

Aprendizaje de Máquina

Tarea 2

Reporte

Integrante:

- Villalón Pineda Luis Enrique

1.- Comparación entre funciones de crecimiento empíricas y teóricas.

- Umbrales (1D): La curva empírica coincide con la teórica $t(m) = m + 1$
- Intervalos (1D): La enumeración exacta iguala

  $t(m) = \frac{m(m+1)}{2} + 1$

  El muestreo aleatorio queda ligeramente abajo por no explorar todos los intervalos.
- Líneas (2D): La teoría es $t(m) = m^2 - m + 2$

  las curvas empíricas van por debajo porque los puntos en grilla generan colinealidades. Con puntos en "posición general" se acercara más.

2.- Pérdida 0–1: el problema de los mínimos locales

Cuando trabajamos con un clasificador simple por umbral en una dimensión, algo interesante ocurre al medir cuántos errores comete para diferentes valores del umbral. La Gráfica 1, del ejercicio 2 nos revela la naturaleza problemática de esta función: la curva de errores tiene forma "escalonada", creando múltiples valles locales donde un algoritmo de optimización podría quedarse atascado. Básicamente si comenzamos la optimización en el lugar equivocado, podríamos terminar en una solución mediocre sin darnos cuenta de que existe una mucho mejor.

2) Pérdida logística: la elegancia de la convexidad

La pérdida logística nos ofrece un panorama completamente diferente. En nuestro experimento, fijamos w=1w = 1 w=1 y variamos únicamente el sesgo para ver cómo se comporta la función de pérdida.

Los resultados en la Gráfica 2 nos dicen que: Forma una curva convexa perfecta con un único mínimo. No importa desde dónde comencemos la optimización todas las trayectorias del descenso de gradiente convergen al mismo punto óptimo, aproximadamente, algo que siempre nos guía hacia la solución correcta.

La Gráfica 3 vemos que mientras la pérdida 0–1 (normalizada) mantiene su estructura escalonada, la logística es suave y presenta un único mínimo bien definido. Por eso es que la regresión logística es tan popular en la práctica: no solo funciona bien, sino que es confiable de optimizar.

3) Cuando el descenso de gradiente se descontrola

Los resultados varían dramáticamente según la tasa elegida:

- $\eta=0.05$: El algoritmo converge de manera suave y predecible (Gráficas 4–5, línea azul). Cada paso nos acerca gradualmente al óptimo.
- $\eta=0.15$: Comenzamos con oscilaciones en la trayectoria (línea naranja), pero el algoritmo eventualmente se estabiliza y converge.
- $\eta=0.25$: El sistema diverge completamente (línea verde), alejándose cada vez más de la solución. Es el equivalente matemático de dar pasos tan grandes que nos alejamos del objetivo con cada intento.

Para esta función cuadrática específica, descubrimos una regla práctica valiosa: mantenemos estabilidad cuando $0<\eta\lambda<2$. Si $\eta\lambda\geq 2$, la divergencia es inevitable.

4) Double descent: cuando más capacidad sorprende

El fenómeno de double descent lo podemos ver como el error de prueba inicialmente baja conforme el modelo se vuelve más complejo, luego sube cerca del punto donde el modelo puede memorizar perfectamente los datos de entrenamiento , pero lo raro es baja otra vez si continuamos aumentando la capacidad.

La interacción entre el modelo elegido, la regularización aplicada y el tamaño de los datos controla este efecto. Es importante notar que aunque la convexidad de la pérdida logística nos garantiza entrenamiento relativamente estable, no previene automáticamente el double descent cuando la capacidad se vuelve grande

5) Lecciones para la práctica

Los hallazgos de este estudio se traducen en recomendaciones concretas:

- $\tau(m)$ es un máximo. Si usas una sola nube de puntos, subestimas; repite con muchas nubes y toma el máximo observado.
- En 1D, empírico ≈ teórico; en 2D, la colocación de puntos determina qué tan cerca estás del máximo.
- La pérdida 0–1 puede tener mínimos locales traicioneros por lo que es difícil de optimizar directamente, mejor usar aproximaciones suaves.
- La pérdida logística es convexa nos ofrece convergencia más confiable y predecible.
- La tasa de aprendizaje es crítica: pequeña significa lento pero seguro; moderada puede variar pero converge; grande lleva a divergenia.
- Para modelos complejos, usa regularización inteligente, validación robusta y controla la capacidad cuidadosamente para navegar los efectos del double descent.

La diferencia entre un modelo que aprende efectivamente y uno que falla es en el hecho de comprender y aplicar estas ideas básicas pero poderosas.