

Apuntes del 27 de febrero del 2024

Anthony Alvarado
Escuela de computación
Instituto Tecnológico de Costa Rica

I. QUIZ #2

- 1) Mencione dos razones del por qué es conveniente escoger MSE en lugar de MAE como loss function.
R/ MAE no es derivable y MSE tiene un mínimo global.

- 2) Mencione que es un mínimo local y un mínimo global en una función.

R/ Un mínimo global es el punto más bajo de una función en todo su dominio, mientras que un mínimo local es el punto más bajo de un vecindario específico de un punto.

- 3) Desarrolle la derivada parcial de L con respecto a W de:

$$\begin{aligned}L &= \frac{1}{N} \sum ((wx_i + b) - y_i)^2 \\ \frac{dL}{dw} &= \frac{1}{N} \sum ((wx_i + b) - y_i)^2 \\ \frac{dL}{dw} &= \frac{1}{N} \sum 2((wx_i + b) - y_i) * ((wx_i + b) - y)' \\ \frac{dL}{dw} &= \frac{1}{N} \sum 2((wx_i + b) - y_i) * ((wx_i + b)' - (y)') \\ \frac{dL}{dw} &= \frac{1}{N} \sum 2((wx_i + b) - y_i) * ((x_i)) \\ \frac{dL}{dw} &= \frac{1}{N} \sum 2((wx_i + b) - y_i) * x_i\end{aligned}$$

Fig. 1.

- 4) Desarrolle la derivada parcial de L con respecto a b:

$$\begin{aligned}L &= \frac{1}{N} \sum ((wx_i + b) - y_i)^2 \\ \frac{dL}{db} &= \frac{1}{N} \sum ((wx_i + b) - y_i)^2 \\ \frac{dL}{db} &= \frac{1}{N} \sum 2((wx_i + b) - y_i) * ((wx_i + b) - y)' \\ \frac{dL}{db} &= \frac{1}{N} \sum 2((wx_i + b) - y_i) * ((wx_i + b)' - (y)') \\ \frac{dL}{db} &= \frac{1}{N} \sum 2((wx_i + b) - y_i) * ((1)) \\ \frac{dL}{db} &= \frac{1}{N} \sum 2((wx_i + b) - y_i)\end{aligned}$$

Fig. 2.

- 5) El learning rate del Descenso del Gradiente es un parámetro o hiper parámetro.

R/ Es un hiper parámetro ya que el valor de alfa es establecido previamente por el ingeniero en vez de ajustarse automáticamente durante el entrenamiento del modelo.

II. SESGO Y VARIANZA

El sesgo es la tendencia de un modelo de machine learning a hacer predicciones inexactas o injustas porque hay errores sistemáticos en el modelo de ML o en los datos utilizados para entrenar el modelo.

La varianza en el aprendizaje automático mide hasta qué punto difieren entre sí los puntos de datos en un conjunto de datos. Cuando la varianza es alta, los puntos de datos tienen una amplia gama de valores y se distribuyen. Un modelo de alta varianza hace que las predicciones estén demasiado alejadas del resultado esperado

Se necesita un modelo que tenga un sesgo bajo y poca varianza.

A. Sesgo alto

- Muchos errores en el training set
- No tiene capacidad para hacer predicciones
- El modelo es simple
- No usa todos los features del modelo

B. Soluciones

- Utilizar un modelo más complejo
- Analizar que features son adecuados y cuales no

C. Varianza alta

- Alta dimensionalidad en los datos
- Posee pocos ejemplos
- Es sensible a las variaciones de los datos
- El modelo no se ajusta a los datos de entrenamiento

D. Soluciones

- Utilizar un modelo más simple
- Reducir la dimensionalidad
- Aplicar técnicas de regularización que hacen el modelo más simple
- Si es posible, obtener más datos para el training set

III. DATASET

Se divide en: training set (conformado por el 80% de los datos) y testing set (conformado por el 20%).

A. *Training set*

- Se usa para ajustar el modelo
- Debería presentar la diversidad de escenarios que se espera

B. *Testing set*

- Se utiliza para evaluar el modelo
- NO se debe utilizar para entrenar el modelo
- Es independiente del training set
- Se calculan métricas para estimar el rendimiento

C. *Posibles escenarios*

- 1) Bajo error en el training set, bajo error en el testing set
 - Este es el escenario ideal y significa que se está utilizando un buen modelo
 - Puede generalizar correctamente
 - El modelo evita que el ruido exista en los datos
- 2) Bajo error en el training set, error alto en el testing set
 - El modelo se ajusta demasiado a los datos (overfitting)
 - No es capaz de generalizar
 - Se obtiene una alta varianza (muchacha diferencia entre el training set y el testing set)
- 3) Error alto en el training set, error alto en el testing set
 - El modelo no aprende nada de los datos (underfitting)
 - Señal de un modelo muy simple
 - Posee un alto sesgo porque el modelo no es capaz de capturar la complejidad de los datos
 - Es peor que el overfitting

REFERENCES

- [1] "Desafío Al Sesgo en el aprendizaje automático: Mailchimp," Mailchimp, <https://mailchimp.com/es/resources/bias-in-machine-learning/#:text=¿Qué%20es%20el%20sesgo%20en,resultado%20de%20su%20valor%20esperado.> (accessed Mar. 1, 2024).
- [2] M. Rouse, "Techopedia: Educating IT professionals to make smarter decisions" Sesgo de las Máquinas, <https://www.techopedia.com/es/definicion/sesgo-maquinas> (accessed Mar. 1, 2024).