

Instituto Tecnológico de Costa Rica
Escuela de Computación

Programa de Maestria en Ciencias de la Computacion

Curso: Aprendizaje Automatico

Profesor: M. Sc. Saúl Calderón Ramírez

QUIZ 0

Entrega: Domingo 13 de Marzo, a través del TEC digital
Debe subir un *pdf* con la respuesta,
junto con el documento lyx, en archivo .zip.

Valor: 100 pts.

Puntos Obtenidos: _____

Nota: _____

Nombre del (la) estudiante: **Andrey Arguedas Espinoza**

Carné: **2020426569**

1. **(70 pts)** Usted y su equipo de machine learning tienen por objetivo implementar un modelo de regresión, para estimar el tiempo en desarrollar un software a partir de la cantidad de desarrolladores y la experiencia promedio del equipo como características, por lo que entonces $\vec{x}_i \in \mathbb{R}^2$. Uno de los miembros del equipo sugiere usar como función de error la función *Mean Squared Logarithmic Error (RMSLE)*. Para optimizar esa función de error, es necesario calcular su vector gradiente respecto a los parámetros del modelo lineal de regresión implementar $y(\vec{x}_i) = \vec{x}_i \cdot \vec{w}$. Calcule el vector gradiente $\nabla_{\vec{w}(t)}$ para tal función de error, para un conjunto de N observaciones $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N$, y sus correspondientes etiquetas $t_i \in \mathbb{R}$:

$$E(\vec{w}(t)) = \frac{1}{N} \sum_{i=1}^N (\ln(y(\vec{x}_i) + 1) - \ln(t_i + 1))$$

donde $\vec{x}_i, \vec{w}(t) \in \mathbb{R}^n$, y $y(\vec{x}_i) = \vec{x}_i \cdot \vec{w}$ corresponde al modelo. Use la regla de la cadena.

RESPUESTA:

1 - Primero sustituimos utilizando $y(\vec{x}_i) = \vec{x}_i \cdot \vec{w}$ por lo que obtenemos :

$$E(\vec{w}(t)) = \frac{1}{N} \sum_{i=1}^N (\ln(\vec{x}_i \cdot \vec{w} + 1) - \ln(t_i + 1))$$

2 - Seguidamente debemos obtener las derivadas parciales para obtener el vector gradiente $\nabla_{\vec{w}(t)}$, necesitamos las derivadas despejando \vec{w} :

Derivada parcial para \vec{w} :

$$\begin{aligned} & \frac{d}{dw} \ln(\vec{x}_i \cdot \vec{w} + 1) - \ln(t_i + 1) \\ &= \frac{d}{dw} [\ln(xw + 1)] + \frac{d}{dw} - \ln(t_i + 1) \\ &= \frac{1}{xw + 1} * \frac{d}{dw} [xw + 1] \end{aligned}$$

$$= \frac{x * \frac{d}{dw}[w] + \frac{d}{dw}[1]}{xw + 1}$$

$$= \frac{x}{xw + 1}$$

Con esto obtenemos la derivada parcial para construir el vector gradiente.

Al llevarlo la derivada a la función de error la función nos quedaría de la siguiente manera:

$$E(\vec{w}(t)) = \frac{1}{N} \sum_{i=1}^N \left(\frac{\vec{x}_i}{(\vec{x}_i * \vec{w}_i) + 1} \right)$$

2. **(30 pts)** Comente, según una muy corta investigación, en no más de 2 párrafos, las ventajas de la función de error RMSLE respecto a la más popular métrica *RMSE* (*Root Mean Squared Error*). Use fuentes bibliográficas académicas y especifíquelas.

RESPUESTA:

Para RMSE los valores atípicos hacen que el resultado de error tome valores muy altos, sin embargo en RMSLE estos son escalados de manera que prácticamente se anula su efecto, por lo que con RMSLE tenemos un valor de error relativo, además RMSLE incurren en una penalización mayor al subestimar y esto puede ser muy útil en casos en los que no es tolerable la sobreestimación pero no hay tanto problema con la sobreestimación.