

# Apuntes Semana 6

Apuntes del 09 de septiembre

Juan Pablo Rodríguez Cano  
IC-6200 Inteligencia Artificial  
Tecnológico de Costa Rica  
jp99@estudiantec.cr

**Abstract**—En este documento se detallan las indicaciones de la tarea 1 de Inteligencia Artificial y se introduce el tema de regresión logística como un modelo de clasificación cuyas propiedades de función son aptas para modelar problemas complejos y la optimización de recursos.

**Index Terms**—

## I. PREGUNTAS DEL QUIZ

- 1) Describa qué es "overfitting" y "underfitting".  
R/ "Overfitting" es cuando el modelo tiene una mejor métrica con el conjunto de entrenamiento que con el conjunto de testing, lo cual indica una pobre generalización con datos nuevos. "Underfitting" es cuando el modelo no logra captar la relación entre los features de manera que los puntajes de métrica son bajos para el conjunto de entrenamiento y testeo.
- 2) Describa k-Fold Cross-Validation  
Se subdivide el conjunto de entrenamiento en k-1 partes. En cada época se entrenan k-1 partes y se utiliza el otro subconjunto para la validación, el
- 3) ¿Qué es un mínimo global y mínimo local en una función?  
Un mínimo local es el valor mínimo de una función en una vecindad reducida, mientras que el mínimo global se refiere al mínimo global a través de todo el dominio de la función.
- 4) Desarrolle la derivada parcial de L con respecto a W de:

$$L = \frac{1}{N} \sum ((wx_i + b) - y_i)$$

$$\frac{\partial L}{\partial w} = \frac{2}{N} \sum ((wx_i + b) - y_i)x_i$$

## II. INDICACIONES DE LA TAREA

- La tarea se debe realizar en grupos de 3 personas.
- La fecha de entrega es el 16 de septiembre.
- Solo hace falta que una persona del grupo suba la tarea. En el nombre del archivo zip debe venir el nombre de todos.
- No se puede utilizar ninguna biblioteca que no sea numpy o pandas
- Kaggle es una plataforma con datasets para machine learning para el público y también presentan oportunidades para participar en concursos de ML.
- La función de pérdida y la graficación debe ser manual

- El método **describe()** resume los datos analíticos que son importantes para saber cómo se comportan los features
- No debe haber código en el informe, solo resultados, análisis etc.
- El notebook será evidencia del trabajo
- El objetivo es ver si la relación con la predicción es lineal, y si no aplicar un feature engineering
- figuras en IEEE siempre van en la parte superior o inferior de las columnas.
- El formato es de IEEE para conferencias

## III. ACTIVIDAD DE IEEE

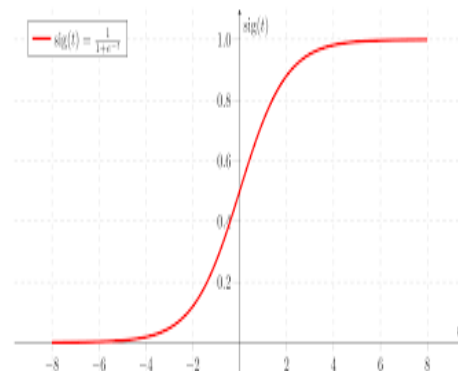
Es un evento anual que se dará esta vez en noviembre en la sabana. Es una oportunidad para conocer sobre temas innovadores en inteligencia artificial y biología molecular. Es una oportunidad para crear contactos dentro de la industria ya que los presentadores suelen ser receptivos al público y disponen de tiempo para hablar.

## IV. CONTENIDO DE CLASE

### A. Regresión Logística

A diferencia de la regresión lineal que es un modelo que predice un número real a partir de los features, la regresión logística es un modelo de clasificación binaria. El resultado de dicho modelo es la probabilidad de que suceda un evento y está basado en la distribución de Bernoulli:  $P(X = k) = p^k(1 - p)^{1-k}$

### B. Función Sigmoide



Esta función es conveniente porque puede modelar comportamientos no lineales, el cual es un comportamiento muy

común en la mayoría de problemas. Trae consigo una mayor complejidad pero a su vez logra resolver problemas más complejos.

Su codominio es de 0 a 1 y esto es muy conveniente ya que los valores probabilísticos comparten ese mismo espacio.

La función sigmoide se expresa de la siguiente manera

$$\sigma(x) = \frac{1}{1 + e^{-x}} \Rightarrow \sigma(f_{w,b}(x)) = \frac{1}{1 + e^{-f_{w,b}x}}$$

La manera en que esta función se convierte en un clasificador es al escoger un umbral. este umbral se utiliza para definir un punto a partir de cuál se califica un evento con una etiqueta o la otra. Por lo general se suele escoger un valor umbral de 0.5.

### C. Derivada de la función Sigmoide

Como la regresión logística es un clasificador, se debe encontrar una función de pérdida adecuada para el problema. Para esto se debe analizar la derivada de la función sigmoide, ya que es necesario para cualquier problema de optimización.

$$\begin{aligned}\sigma'(x) &= \frac{1'(1 + e^{-x}) - (1(1 + e^{-x}))'}{(1 + e^{-x})^2} \\ \Rightarrow \sigma'(x) &= \sigma(x)(1 - \sigma(x))\end{aligned}$$

Como se puede notar, la derivada se puede expresar en términos de la función misma, lo cual lo hace muy conveniente ya que no se requieren operaciones muy complejas y con esto se obtiene una mayor eficiencia.

### D. Función de Pérdida: Verosimilitud

En vez de utilizar MSE o MAE, se utiliza la verosimilitud. Esta está dada por la siguiente ecuación

$$L = \prod f_{w,b}(x_i)^{y_i} (1 - f_{w,b}(x_i))^{1-y_i}$$

El resultado que se obtiene para un punto en esta ecuación es la probabilidad de que su etiqueta sea  $y_i$  con los pesos  $w$  actuales. Como se quiere optimizar los pesos para los cuales se obtiene una mejor métrica, se debe derivar esta función. Sin embargo, existe un problema con esta expresión donde una multiplicación incluye polinomios muy grandes, y calcular la derivada respectiva se vuelve muy complejo y computacionalmente costoso. Además, como se trata de valores probabilísticos, o sea, de 0 a 1, su multiplicación se vuelve extremadamente pequeña y así la derivada de la función se vuelve virtualmente cero, y esto no cambia los pesos en el paso de entrenamiento. A esto se le conoce como el fenómeno de "vanishing gradients". Por esta razón se aplican los teoremas de logaritmo y se obtiene la siguiente expresión.

$$\begin{aligned}\ln(L) &= \sum \ln(f_{w,b}(x_i)^{y_i}) + \ln((1 - f_{w,b}(x_i))^{1-y_i}) \\ \Rightarrow \ln(L) &= \sum y_i \ln(f_{w,b}(x_i)) + (1 - y_i) \ln((1 - f_{w,b}(x_i)))\end{aligned}$$

Esto se convierte en una tarea más fácil de optimización. Sin embargo, la función de logaritmo es estrictamente creciente,

por lo que hay que convertir un problema de maximización en minimización. Para esto, simplemente se da vuelta a la función de  $\ln$  multiplicando por -1.