

# Laboratorio 3: Naive Bayes, SVM y Árboles

## Inteligencia Artificial - CC3085

José Antonio Mérida Castejón

8 de febrero de 2026

### Task 1 - Teoría

Responda con criterio y análisis de ingeniero. No se esperan definiciones de libro, si no que analice las consecuencias de las decisiones de diseño.

#### La Mentira de la Independencia: Naive Bayes

*En la diapositiva 6 se menciona que Naive Bayes “Asume independencia entre observaciones”. Con esto en mente, considere que usted está construyendo un filtro de spam y su modelo analiza la frase “Cuenta Bancaria”. En el mundo real, la probabilidad de que aparezca la palabra “Bancaria” aumenta drásticamente si ya apareció la palabra “Cuenta”. Responda:*

- *Si Naive Bayes trata estas dos palabras como eventos independientes (lanzar dos monedas separadas) en lugar de eventos dependientes, ¿está el modelo subestimando o sobreestimando la probabilidad real de la frase conjunta? Justifique su respuesta basándose en la fórmula  $P(A \cap B) = P(A) \times P(B)$  vs la realidad.*

Naive Bayes, al asumir que los eventos son independientes es completamente incapaz de determinar relaciones entre ellos. Desde un aspecto matemático, podemos decir que  $P(B | A) > P(B)$  dado el enunciado que establece que la probabilidad que aparezca “Bancaria” aumenta al aparecer “Cuenta”. Por lo tanto, el modelado más adecuado (bajo la observación dada en el inciso) nos indica que la probabilidad real es  $P(A \cap B) = P(A) \times P(B | A)$ . Mientras que NB nos dice “*estos eventos son completamente independientes, entonces la probabilidad que ocurra B luego de A es  $P(A) \times P(B)$* ”.

Como establecimos previamente,  $P(B | A) > P(B) \implies P(A) \times P(B | A) > P(A) \times P(B)$ . Es decir, nuestro modelo subestimaría la probabilidad de la frase real conjunta. Esta problemática también pone en duda el desempeño de Naive Bayes en cuanto a modelos que involucran el procesamiento de lenguaje natural. Existen algunas relaciones como frases conjuntas, o *secuencias específicas* que proporcionan significado más allá de las palabras individuales.

#### La Economía de los Datos (SVM)

*Refierase a las slides correspondientes al tema de SVM. Con esto en mente considere, usted entrena un SVM con 1 millón de datos de partidas de League of Legends. El modelo resultante identifica 5,000 “Vectores de Soporte”. Su jefe le dice que para ahorrar espacio en la base de datos, va a eliminar los otros 995,000 puntos de datos que no son vectores de soporte y re-entrenar el modelo solo con los 5,000 restantes. Responda:*

- *Matemáticamente, cambiaría la frontera de decisión al hacer esto? Por que si o por que no?*

No, la frontera sería la misma ya que esta se determina únicamente por los vectores de soporte. Analizando un poco más a profundidad, tenemos que el vector  $w$  que define los pesos es el siguiente:

$$\sum_{i=1}^n \alpha_i y_i x_i$$

Donde  $\alpha_i$  representa el multiplicador de Lagrange de cada punto  $x_i$ . Entonces, por definición sabemos que únicamente los vectores de soporte tienen  $\alpha_i > 0$  y son los únicos que tienen efecto sobre  $w$ .

- *Explique la eficiencia de memoria de este algoritmo frente a KNN.*

Este algoritmo es más eficiente en memoria, en este caso por ejemplo si tuviésemos un KNN deberíamos de mantener el total de 1 millón de puntos en memoria. Mientras que en SVM únicamente debemos mantener los 5,000 vectores de soporte mencionados.

**La Miopía de los Árboles** *En las diapositivas se menciona que la construcción del árbol es un “algoritmo greedy” (codicioso /avaro). Con esto en mente considere, un algoritmo greedy toma la mejor decisión posible en el paso actual sin preocuparse por el futuro. Responda:*

- *Si el árbol elige el “Mejor Feature” para dividir el nodo raíz porque reduce la impureza drásticamente ahora, ¿garantiza esto que el árbol final será el más óptimo/pequeño posible?*

No necesariamente, una división “mala” puede permitirnos realizar divisiones perfectas en los siguientes niveles. Al utilizar un approach greedy, descartamos completamente las posibilidades a futuro y elegimos el beneficio “local” por encima de la solución global. Tomando como ejemplo un juego de ajedrez, existen jugadas donde el sacrificio de una pieza resulta en una mejor posición algunos movimientos después. Si tomamos un approach greedy, solo estaríamos viendo la pérdida inmediata de material y descartaríamos el movimiento completamente.

- *Dibuje o describa un escenario lógico donde elegir una división “sub-óptima” al inicio podría llevar a un mejor árbol al final. Qué nombre técnico recibe este fenómeno*

Supongamos que tenemos un dataset y estamos entrenando un modelo inteligente para saber cuando se abre una bóveda en un banco. Esta bóveda requiere de dos llaves, de manera que si ambas son giradas se abre la bóveda. Sin embargo, tenemos una variable adicional que representa si la luz del pasillo estaba encendida o no (completamente irrelevante para la apertura de la bóveda). Claramente, el modelo óptimo ignoraría por completamente la variable adicional de luz encendida y sería una simple decisión para encontrar  $llave_1 = 1, llave_2 = 1$  donde de lo contrario se encuentra cerrada la bóveda. Sin embargo, al momento de dar el primer paso puede que el modelo encuentre lo siguiente:

- $llave_1$ : Como ambas llaves necesitan girarse, el algoritmo nota que aunque esta este girada la mayoría de los casos no indican que la bóveda se abra.
- $llave_2$ : De manera idéntica al anterior, el algoritmo ve que la llave se gira pero no se correlaciona directamente con la apertura de la bóveda.

- *luz*: Gracias al ruido estadístico, el 60 % de observaciones dónde la luz está encendida la bóveda está abierta.

Tomando un approach greedy, claramente la decisión inicial es separar los datos con la variable *luz*. Mientras que por medio de una decisión subóptima (como elegir *llave*<sub>1</sub> inicialmente, a pesar que no nos beneficie de momento) nos permite realizar las divisiones necesarias para encontrar el único caso dónde se abre la bóveda (ambas llaves giradas). Es decir, puede que algunas variables por si solas no sean beneficiosas de momento, pero esto no significa que sean malas elecciones.

A este fenómeno se le conoce como *Greedy Myopia*.