Probabilidad y estadística para la Inteligencia Artificial

Camilo Enrique Argoty Pulido

Especialización en Inteligencia Artificial

5 de marzo de 2025

Laboratorio de Sistemas Embebidos

Plan del curso

El objetivo del curso es dar las herramientas de estadística y probabilidad básicas y necesarias para el trabajo en aprendizaje de máquina e inteligencia artificial.

El plan del curso es el siguiente:

- Clase 1: Probabilidad, enfoque frecuentista y bayesiano. Espacios, eventos, probabilidad condicional, regla de Bayes, ley de la probabilidad total e independencia.
- Clase 2: Variables aleatorias discretas y continuas, esperanza, varianza, función de distribución, función de densidad, funciones empíricas, aproximación de densidad en Python y distribuciones importantes.
- Clase 3: Estimadores de mínimos cuadrados y máxima verosimilitud.
- Clase 4: Teorema de los grandes números, teorema central del límite y distribución normal. Intervalos de confianza.
- Clase 5: Inferencia estadística, pruebas de hipótesis para la media.
- Clase 6: Pruebas de hipótesis para la media con pocos datos, varianza, diferencia de medias, ANOVA.
- Clase 7: Inferencia bayesiana.
- Clase 8: Repaso y preguntas.

Evaluación

La evaluación del curso consiste en 3 trabajos prácticos, con valor de 30 %, 30 % y 40 %.

La nota final es el promedio ponderado de las notas de los trabajos prácticos con los pesos mencionados.

Bibliografía Recomendada:

Este curso se encuentra en prácticamente autocontenido, donde las grabaciones y las presentaciones serán accesibles para los estudiantes. Sin embargo, los siguientes libros son fuentes complementarias recomendables:

- Walpole R. Myers R. Myers S. Ye K. : Probabilidad y estadística para ciencias e ingeniería.
- Jackman S. : Bayesian analysis for social Science.

Probabilidad

Contexto y necesidad histórica

Lo que hoy se conoce como probabilidad, surge de la inquietud de por tener mejor desempeño en los juegos de azar.

Algunos de estos personajes fueron Girolamo Cardano, Pierre de Fermat, Blaise Pascal y Christian Huygens.

En búsqueda de un número predictor

Para aquellos pioneros de la probabilidad, pronto quedó clara la necesidad de un número que permitiera comparar la 'factibilidad' de los eventos, en el sentido de saber entre dos eventos, cuál era más fácil que ocurriera.

De aquí viene la palabra *probabilidad* como qué tan fácil es probar un evento, cuando probar se entiende como evidenciar que un evento ocurre.

Similitud con los conceptos de área y volumen

De lo anterior viene la pregunta, ¿Qué se entiende por evento?

Pronto fue claro que el número que se buscaba tenía que ser una función de conjunto, en vez de ser una función puntual, es decir, debía depender de subconjuntos de un conjunto más grande, denominado *espacio muestral*.

Esto lleva a entender la probabilidad como un concepto muy parecido al de área o volumen.

Aproximación frecuentista

De las primeros resultados que se obtuvieron, fue la famosa *Fórmula de Laplace:*

$$P(A) = \frac{(Numero\ de\ casos\ favorables)}{(Numero\ de\ casos\ posibles)}$$

Esto dió lugar al denominado *Enfoque Frecuentista* de la probabilidad, en el cual toda probabilidad se interpreta como el resultado asintótico a largo plazo de repetir un experimento.

Existe otro enfoque, denominado *Enfoque Bayesiano*, en el cual la probabilidad es la creencia subjetiva sobre la ocurrencia de un evento. Dicha creencia puede modificarse a medida que tenemos acceso a más y más datos.

Esto tiene que ver con el concepto de *probabilidad condicional*

Probabilidad condicional

Pronto Thomas Bayes llegaría a lo que se conoce como la *probabilidad condicional de un evento A, dado un evento B:*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

De aquí también viene el concepto de **independencia**. Dos eventos se dicen

independientes si

$$P(A|B) = P(A)$$

Es decir, que el evento B no da información sobre el evento A.

De forma equivalente, A y B son independientes si:

$$P(A \cap B) = P(A)P(B)$$

Ley de probabilidad total

Esto lleva a la *Ley de probabilidad total* Si $(B_i)_{i\in I}$ es un conjunto de eventos dos a dos disyuntos cuya unión es el espacio muestral Ω, y A es un evento,

$$P(A) = \sum_i P(A|B_i)P(B_i),$$

Teorema de Bayes

Si $(B_i)_{i\in I}$ es un conjunto de eventos dos a dos disyuntos cuya unión es el espacio muestral Ω, y A es un evento, dado un $j \in I$

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_i P(A|B_i)P(B_i)},$$