
TEMARIO

Estadística inferencial

Descripción

El fundamento de una gran mayoría de las técnicas avanzadas de análisis de datos se encuentra en la estadística. Por ejemplo, los llamados variational autoencoders son redes neuronales profundas usadas como modelos generativos y que contienen importantes bases probabilísticas y estadísticas. Por ejemplo, dentro de una de sus capas incluyen parámetros para la media y la dispersión. Ni que decir de muchos modelos en machine learning, como la regresión logística o el análisis discriminante, los cuales nacieron como modelos en esa área. Incluso técnicas indispensables en el preprocesamiento de información, como la estandarización, las técnicas de manejo de datos perdidos, así como la selección de variables, requieren amplios conocimientos en Estadística. Incluso si nuestro interés es solo predictivo, el entendimiento de porque las predicciones tienen cierta variabilidad y con mayor razón, determinar a cuál corresponde, se fundamenta en conceptos estadísticos. Por lo tanto, es difícil pretender tener una carrera completa en Ciencia de Datos si nos faltan los cimientos indispensables para ello dados en la Estadística.

El objetivo de este curso es introducir al estudiante en el área de la estadística. Se parte desde el análisis exploratorio de datos, parte importante para entender tanto numéricamente como gráficamente a un conjunto de datos. Posteriormente, se determinan estimadores puntuales asociados a los parámetros de una distribución, por ejemplo, la media en una distribución normal. Estos son funciones de los valores asociados a una muestra aleatoria y que por lo tanto poseen cierta variabilidad. Por lo tanto, uno se pregunta si cualquier función de la muestra o estimador es igual de bueno para estimar un parámetro. Este cuestionamiento hace necesario el plantear cuáles propiedades debería de cumplir un estimador puntual para considerarse bueno. Dada la variabilidad que un estimador posee, uno se cuestiona también si no es más conveniente dar un rango de posibles valores que puede tomar el parámetro en lugar de solo un valor, esta es la estimación intervalar. Finalmente, uno puede preguntarse si nuestra noción de que un parámetro según nuestro conocimiento debe de tomar un valor o valores específicos es respaldada por nuestros datos, las pruebas de hipótesis permiten realizar este paso.

1. Análisis exploratorio de datos (AED)
 - a) Análisis de frecuencias
 - (i) Moda

- b) Representación gráfica de las frecuencias
 - c) Métodos de análisis de frecuencias adicionales para variables cuantitativas
 - (i) Análisis gráficos
 - (ii) Medidas de tendencia central y de dispersión
 - (iii) Resumen de los cinco números
 - d) Comparación de poblaciones
 - e) Asociación entre variables
 - (i) Correlación de Pearson y gráfica de dispersión
 - (ii) Asociación entre variables cualitativas
2. Estimación puntual
- a) Muestra aleatoria y estadísticos
 - b) Estimación puntual
 - c) Métodos para encontrar estimadores
 - (i) Método de momentos
 - (ii) Método de máxima verosimilitud
 - d) Propiedades de los estimadores puntuales
 - (i) Sesgo de un estimador puntual
 - (ii) Error Cuadrático Medio
 - (iii) Consistencia
 - (iv) Familia exponencial de densidades
 - (v) Familia exponencial k-dimensional
 - (vi) Suficiencia
 - (vii) Eficiencia y completez
 - (viii) Propiedades asintóticas de los estimadores de máxima verosimilitud
 - e) Propiedades de los estimadores de una m.a. normal
3. Intervalos de confianza
- a) Método pivotal
 - b) Intervalos de confianza para una distribución Normal
 - (i) Intervalo de confianza para la media
 - (ii) Intervalo de confianza para la varianza
 - (iii) Intervalo de confianza para la diferencia de medias de dos poblaciones normales
 - (iv) Intervalo de confianza para el cociente de varianzas cuando las medias son desconocidas
 - c) Intervalos de confianza aproximados en proporciones
 - (i) Intervalos de confianza aproximados para proporciones
 - (ii) Intervalos de confianza aproximados para diferencias de proporciones.
 - d) Intervalos de confianza para muestras no normales.
4. Pruebas de Hipótesis
- a) Introducción

- b) Hipótesis estadística
- c) Tipos de errores
- d) Función potencia
- e) Tamaño de la prueba
- f) Hipótesis simples
 - (i) Prueba más potente
 - (ii) Lema de Neyman-Pearson
- g) Hipótesis compuestas
 - (i) Cociente de verosimilitudes
 - (ii) Prueba uniformemente más potente
- h) Relación entre intervalos de confianza y pruebas de hipótesis
- i) Pruebas de hipótesis que involucran a la distribución normal
- j) El p valor o nivel crítico

Informes: lumialearning@gmail.com