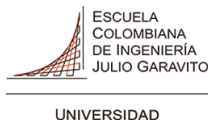


DataVR

Miguel Ángel Hernández Vargas y Diego Andrés Avella Acevedo

Ingeniería Estadística
Escuela Colombiana de Ingeniería Julio Garavito

Noviembre, 2024



- ① Introducción
- ② Objetivos
- ③ Análisis exploratorio
- ④ Regresión logística
- ⑤ Visualización en Realidad Virtual
- ⑥ Futuros desarrollos

- 1 Introducción
- 2 Objetivos
- 3 Análisis exploratorio
- 4 Regresión logística
- 5 Visualización en Realidad Virtual
- 6 Futuros desarrollos

Introducción



- 1 Introducción
- 2 Objetivos**
- 3 Análisis exploratorio
- 4 Regresión logística
- 5 Visualización en Realidad Virtual
- 6 Futuros desarrollos

Objetivos

Objetivo general

- Desarrollar una herramienta interactiva en realidad virtual que facilite el análisis descriptivo y predictivo de datos, integrando técnicas como el análisis de componentes principales (PCA) y la regresión logística, para mejorar la clasificación de datos y detectar posibles errores de etiquetado en un conjunto de datos.

Objetivos

Objetivos específicos

- Implementar un modelo de regresión logística que clasifique un conjunto de datos en función de una variable de respuesta, evaluando su rendimiento mediante métricas estándar como la precisión, la matriz de confusión, y el AIC.
- Mostrar los resultados del entrenamiento del modelo, incluyendo las predicciones y las clasificaciones reales, para permitir la comparación visual y la identificación de posibles errores de clasificación.
- Desarrollar una funcionalidad para identificar y resaltar puntos en los datos que podrían estar mal etiquetados, comparando la clasificación predicha por el modelo con la categoría original de los datos.

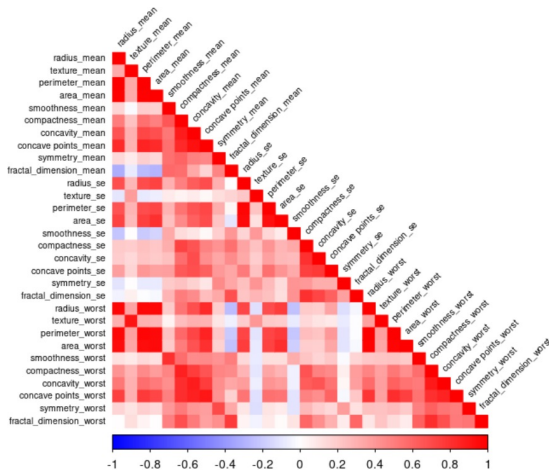
- 1 Introducción
- 2 Objetivos
- 3 Análisis exploratorio**
- 4 Regresión logística
- 5 Visualización en Realidad Virtual
- 6 Futuros desarrollos

EDA

Variables

- Radius
- Texture
- Perimeter
- Area
- Smoothness
- Compactness
- Concavity
- Concave points
- Symmetry
- Fractal dimension

Matriz de correlación



- 1 Introducción
- 2 Objetivos
- 3 Análisis exploratorio
- 4 Regresión logística**
- 5 Visualización en Realidad Virtual
- 6 Futuros desarrollos

Regresión Logística

- **Objetivo:** Clasificación binaria de variables dependientes.
- Utiliza la **función sigmoide** para mapear las predicciones a probabilidades entre 0 y 1.
- **Modelo:**
$$P(y = 1|X) = \frac{1}{1+e^{-(\beta_0+\beta_1X_1+\dots+\beta_nX_n)}}$$
- **Evaluación:** Se mide mediante precisión, matriz de confusión, AIC, y curvas ROC.
- **Ventaja:** No requiere supuestos distribucionales sobre las variables independientes.

- 1 Introducción
- 2 Objetivos
- 3 Análisis exploratorio
- 4 Regresión logística
- 5 Visualización en Realidad Virtual**
- 6 Futuros desarrollos

VR

Pasamos al entorno VR

- 1 Introducción
- 2 Objetivos
- 3 Análisis exploratorio
- 4 Regresión logística
- 5 Visualización en Realidad Virtual
- 6 Futuros desarrollos**

Futuros desarrollos

- **Selección de variables para panel EDA:** Refactorizar el panel que contiene el resumen estadístico básico del conjunto de datos a un dashboard que permita seleccionar las variables y el tipo de gráfico que se desea realizar para su análisis.
- **Incorporar otros modelos de clasificación:** Extender las herramientas de clasificación, utilizando más modelos como árboles de decisión, LDA o QDA, con el fin de comparar su rendimiento entre sí y ofrecer al usuario una visión más amplia sobre qué modelo funciona mejor con los datos utilizados.
- **Implementación de predicciones en tiempo real:** Permitir que el modelo realice predicciones en tiempo real, es decir, que los usuarios puedan cargar nuevos datos y ver instantáneamente la predicción y su probabilidad.

Gracias