

CENTRO DE INVESTIGACIÓN EN
COMPUTACIÓN

CLASIFICACIÓN INTELIGENTE DE PATRONES

Representación de datos

Autor:
Diego Noguez Ruiz

Profesor:
Dr. Cornelio Yáñez
Márquez

1 de Abril 2023



Índice general

Índice general	1
1. Introducción	3
2. Desarrollo	3
2.1. Propósito de la tarea	3
2.2. Instrucciones	3
2.3. Parte 1	3
2.4. Gráficas bidimensionales	3
3. Conclusiones y trabajo futuro	5

Índice de figuras

Figura 1 - Scatterplot Embarazos vs Edad	4
Figura 2 - Scatterplot Glucosa vs Presión Arterial	4
Figura 3 - Scatterplot Grosor de la piel vs Insulina	5

1. Introducción

La vista es uno de los sentidos más importantes del ser humano. De hecho, el cerebro realiza automáticamente una gran cantidad de clasificaciones a través de este sentido. Es por ello que representar datos de manera visual es crucial, ya que nos proporciona una perspectiva diferente de la información que se tiene. Por esta razón, en la última sesión aprendimos a representar datos en 2 y 3 dimensiones, utilizando colores para distinguir entre distintas clases.

En esta tarea, realizaremos el cálculo correspondiente de la cantidad de gráficos distintos que se pueden generar en el plano, y pondremos en práctica lo aprendido.

2. Desarrollo

2.1. Propósito de la tarea

Calcular y representar los valores de las celdas en las matrices de confusión para problemas biclase.

2.2. Instrucciones

De acuerdo con la expresión de la diapositiva 17, ¿cuántas gráficas en el plano cartesiano (vistas bidimensionales) se obtienen con las dos clases del dataset pima de KEEL? Presentar y explicar tres de esas gráficas bidimensionales.

2.3. Parte 1

Pregunta: ¿Cuántas gráficas en el plano cartesiano (vistas bidimensionales) se obtienen con las dos clases del dataset pima de KEEL?

El data set Pima [1] cuenta con 8 atributos, por lo tanto se pueden realizar un total de $\binom{8}{2} = 28$ combinaciones distintas

2.4. Gráficas bidimensionales

Para todas de estos plots, aquellos puntos de color rojo, representan a la clase positiva a diabetes, por otro lado, los puntos de color azul representan a la clase negativa a diabetes. Se puede pensar que cada uno de estos plots da 1/28 de información total.

- Embarazos vs Edad

En este scatter plot el eje de las abscisas queda representado por la edad y el eje de las ordenadas las embarazadas. Estos dos atributos planteados de esta manera no ayuda a discernir muy bien entre clases. Véase figura 1

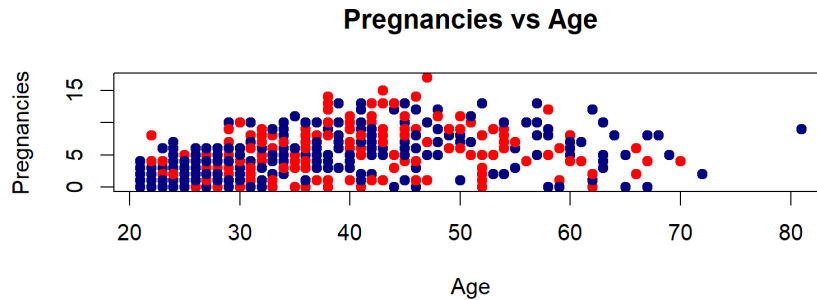


Figura 1: Scatterplot Embarazos vs Edad

- Glucosa vs Presión Arterial

En este scatter plot el eje de las abscisas queda representado por la glucosa y el eje de las ordenadas por la presión arterial. En este plot se alcanza a ver un poco la separación entre clases, podemos notar que aquellos con diabetes tienen una glucosa más alta en comparación con los sanos. Véase figura 2

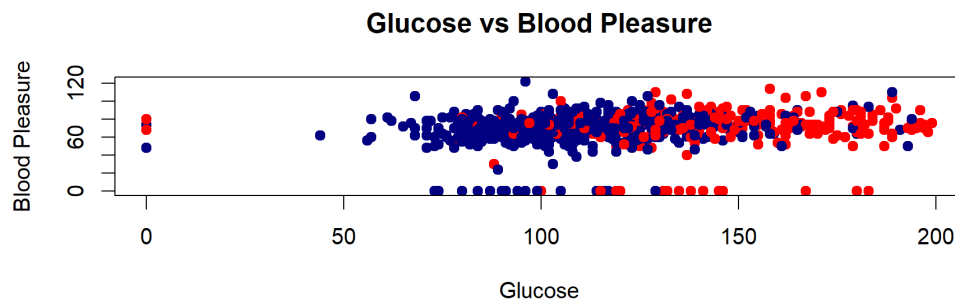


Figura 2: Scatterplot Glucosa vs Presión Arterial

- Grosor de la piel vs Insulina

En este scatter plot el eje de las abscisas queda representado por la el grosor de la piel y el eje de las ordenadas por la insulina. En este plot todos los puntos se encuentran muy juntos, por lo cual es difícil poder sacar alguna conclusión solamente usando esto. Véase figura 3

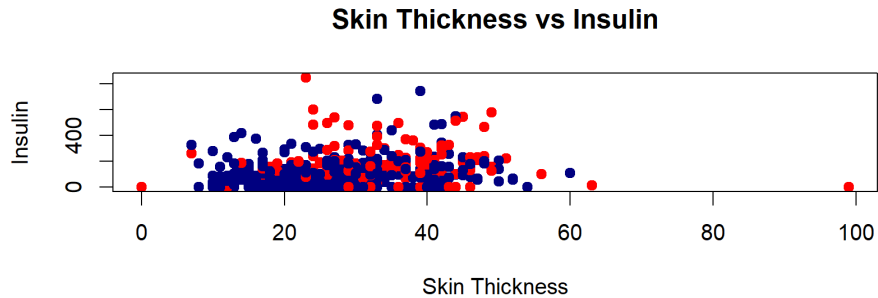


Figura 3: Scatterplot Grosor de la piel vs Insulina

3. Conclusiones y trabajo futuro

Con esta tarea, he llegado a la conclusión de que los gráficos bidimensionales proporcionan poca información sobre las clases debido a dos. En primer lugar, el número de atributos influye en la fracción de información que queda representada en el gráfico, ya que a medida que aumenta la cantidad de atributos, el porcentaje de información obtenida igual disminuye. En segundo lugar, creo que para obtener una información más precisa a través de los gráficos, es necesario seleccionar características que sean distintas entre sí, ya que algunos atributos pueden ser una combinación lineal de otros, lo que dificulta la distinción entre clases.

Por lo tanto, considero que es importante explorar otras alternativas para visualizar datos que permitan mejorar estas limitaciones en el futuro, e infiero que en breve las aprenderemos.

Referencias

- [1] National Institute of Diabetes and Digestive and Kidney Diseases//Kaggle, “Pima indians diabetes database.” <https://sci2s.ugr.es/keel/dataset.php?cod=155>. Visitado en Marzo 2023.