Tarea 1

Estudiantes

John Daniel hoyos Arias
Ivan Santiago Rojas Martinez
Genaro Aristizabal

Docente

Juan Carlos Salazar Uribe

Asignatura

Analitica de datos

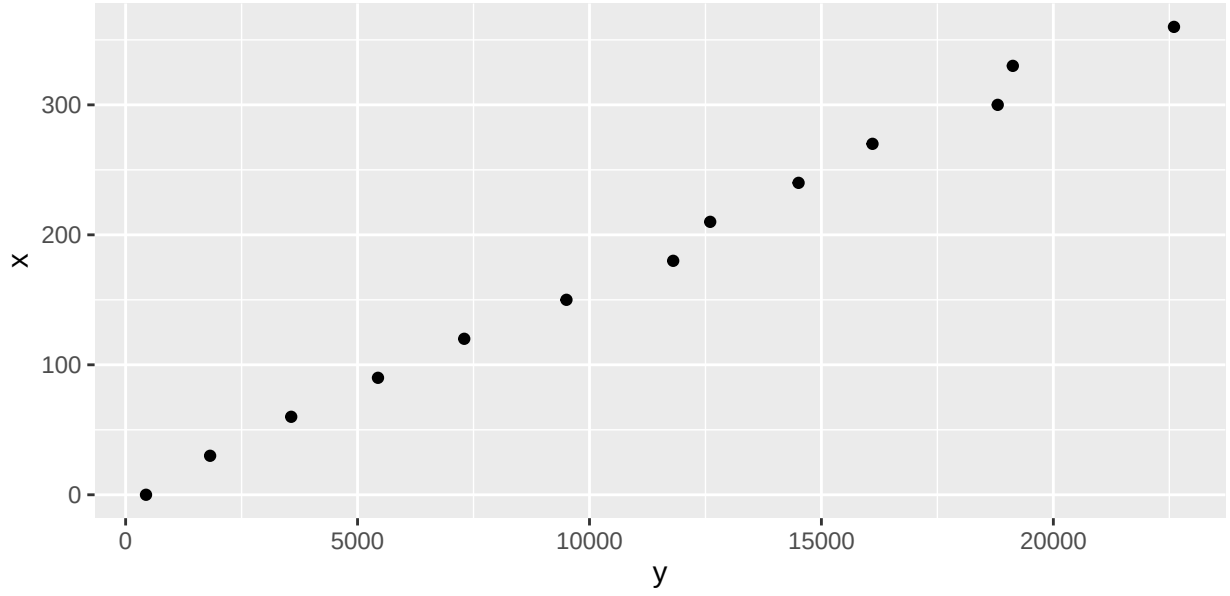
Sede Medellín
17 de septiembre del 2022

Índice

Índice de figuras

1. Ejercicio1

2. Ejercicio2



3. Ejercicio3

3.1. K-nearest neighbors (KNN)

$$Pr(Y = J \mid X = x_0) \approx \frac{1}{K} \sum_{i \in N_0} I(y_i = j)$$

Cuadro 1: Base de datos

X1	X2	X3	Y
0	3	0	Red
2	0	0	Red
0	1	3	Red
0	1	2	Green
-1	0	1	Green
1	1	1	Red

3.2. a) Distancia a cada observación

Usando la distancia euclidiana entre dos punto u y v definida como:

$$d(u,v) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 - v_3)^2}$$

Calculamos la distancia entre cada observación y el punto $X_1 = X_2 = X_3 = 0$ usando R.

```
point <- c(0, 0, 0)

dist_eucl <- function(x){
  ans <- c()
  for (i in 1:nrow(x)){
    xi <- as.numeric(t(as.vector(x[i, ])))
    result <- sqrt(sum((xi-point)^2))
    ans <- append(ans, result)
  }
  return (ans)
}

db <- mutate(db, dist = dist_eucl(db[1:3]))
```

Se obtiene:

Cuadro 2: Distancia a cada observación desde el punto $X_1 = X_2 = X_3 = 0$

Observación	Grupo	Distancia Euclidiana
1	Red	3.000000
2	Red	2.000000
3	Red	3.162278
4	Green	2.236068
5	Green	1.414214
6	Red	1.732051

3.3. b) Predicción para K = 1

Con una selección de K = 1. Knn identifica la observación más cercana al punto con características $X_1 = X_2 = X_3 = 0$ y en este caso la observación mas cercana es la **numero 5** con una distancia de **1.414214**. Dando así Knn una estimación de 1/1 de pertenecer al grupo **Green**. Por ende la estimación es pertenecer a la clase **Green**.

3.4. c) Predicción para K = 3

Con una selección de K = 3. Knn identifica las 3 observaciones más cercanas al punto con características $X_1 = X_2 = X_3 = 0$ y en este caso las observaciones mas cercana son la **numero 5**, la **numero 6** y la **numero 2** que consisten en 2 observaciones de la clase **Red** y una observación de la clase **Green**, dando como restulado una estimación de 2/3 de pertenecer a la clase **Red** y 1/3 de pertenecer a la clase **Green**. Por consiguiente se estima pertenecer a la clase **Red**.

3.5. d) Frontera de decisión de Bayes

Si la frontera de decisión de Bayes en este problema es altamente no lineal, ¿esperaríamos que el mejor valor de K fuera grande o pequeño? ¿Por qué?

4. Ejercicio4