

# Tarea 2

Francisco Alonso

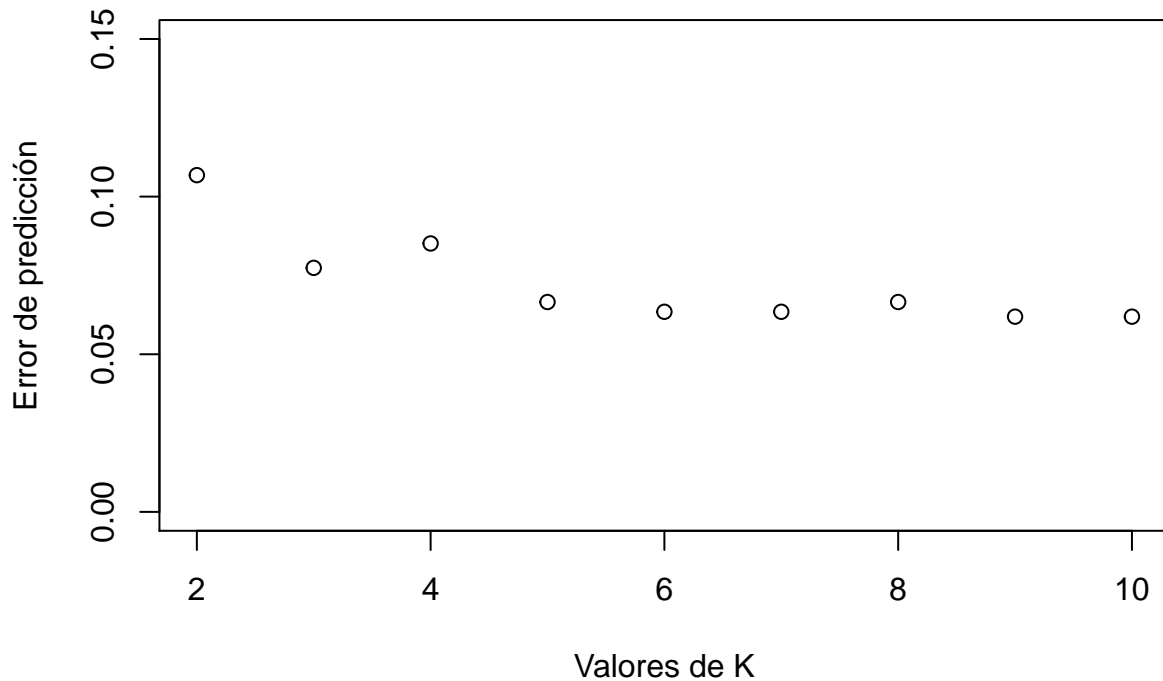
April 27, 2016

## Ejercicio 1

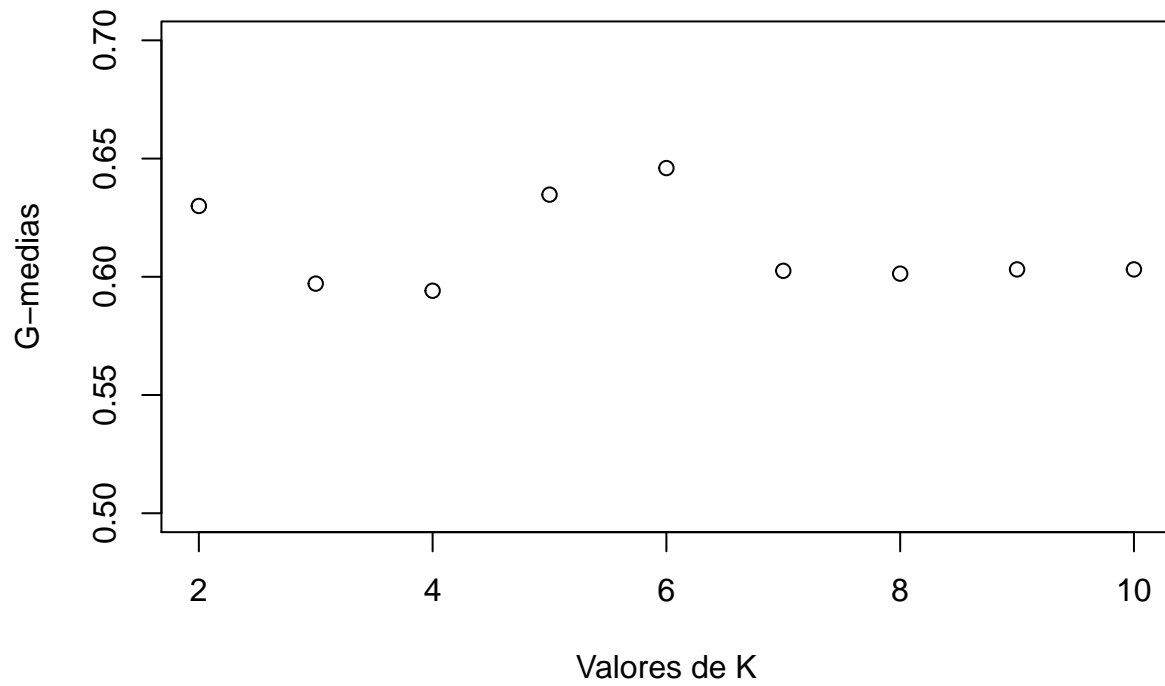
Se aplica el proceso de minería de de datos sobre el conjunto de datos “SEISMIC-BUMPS”, en este están clasificados eventos en minas según si en el siguiente estado ocurrió o no un movimiento sísmico con mucha energía, este se clasifica como “estado peligroso” (1) o “estado no peligroso” (0) según un conjunto de 13 variables correspondientes a mediciones de energía en movimientos sísmicos y otras 5 que indican, en general, el nivel de peligrosidad de los movimientos ya ocurridos.

Se prepara la data para la creación del modelo, es necesario que las variables nominales sean convertidas a valores numéricos, luego se divide el conjunto de datos en un conjunto de entrenamiento y uno de prueba para construir nueve modelos usando el algoritmo k-vecinos más cercanos (k nearest neighbors - knn) variando el número de grupos de dos a diez. Por último se toman medidas del error de predicción y g-medias para comparar cada modelo construido.

### Error de predicción para cada valor K



### G-medias para cada valor K



Se observa que el error de predicción se reduce al usar  $k = 9$ , sin embargo dado que esta métrica es susceptible al sesgo entre clases y observamos que las clases no están bien representadas en el conjunto de datos, nos referimos al valor de g-medias para determinar que el valor de  $k$  apropiado es cuatro.

```
##      0      1
## 2414  170
```

