COMP 6315

Tarea 2

Fecha de entregar: Jueves, Octubre 4, 2018, hasta las 8pm

**Datasets: Estan disponibles en <https://archive.ics.uci.edu/ml/datasets.html>**

PI y P II: Default of credit cards

PIII y PIV: Wearable Computing: Classification of Body Postures and Movements (PUC-Rio) (Anibal,Nitza, Cristian)

PIII y PIV: Online News Popularity Data Set (Claudio, Jared, Ricardo)

PIII y PIV: Activity recognition with healthy older people using a batteryless wearable sensor Data Set (Michelle, Francisco,Lesbia)

**Usar R (dprep u otras librerias) , Python o Rapidminer.**

**Los errores de clasificación deben ser estimados usando validación cruzada O por el Holdout método.**

I(10). Aplicar el metodo RELIEF a su conjunto de datos para seleccionar el mejor subconjunto de variables. Tratar dos distintos valores como umbrales. Evaluar el subconjunto elegido usando el estimado del error de clasificación del clasificador LDA o naive Bayes.

II(8). Discretizar su conjunto de datos usando cualquier método y calcular la inconsistencia de su conjunto de datos discretizado y seleccionar los mejores subconjuntos de variables usando el metodo LVF (inconsistencia y seleccion random) o FINCO (inconsistencia y forward selection).

III(8). Aplicar un metodos wrapper, junto los clasificadores LDA y naive Bayes, para seleccionar variables en su conjunto de datos y comparar sus resultados con las partes I y II.

IV(14) Aplicar componentes principales (PCA) a su conjuntos de datos y luego usar el clasificador LDA para predecir las clases usando los mejores PC. Estimar el error de clasificación usando LDA y/o Naive Bayes. Aplicar nuevamente PCA al conjunto de datos original pero usando solo las mejores variables seleccionadas( I, II o III). Estimar nuevamente el error de clasificación y compararlo con el resultado anterior.