

# Teorías de información – 543727

## Miniproyecto N° 2

Jorge E. Pezoa & Sebastian E. Godoy  
[jpezoa@udec.cl](mailto:jpezoa@udec.cl), [segodoy@udec.cl](mailto:segodoy@udec.cl)

31 de octubre de 2018

Este miniproyecto se debe trabajar en grupos y debe entregarse a más tardar el **miércoles 14/11/2018 a las 23:59 hrs.** vía email a [jpezoa@udec.cl](mailto:jpezoa@udec.cl).

### Clasificación lineal, Bayesiana y Neyman-Pearson

Utilizando el archivo `SkillCraft1_Dataset.csv`, se pide diseñar clasificadores lineales, bayesianos y de Neyman-Pearson. De aquí en adelante se considerarán dos problemas de clasificación: (i) binario donde los League Indexes del 1 al 5 representan la clase 1 y los League Indexes 6, 7 y 8 representan la clase 2; (ii) ternario donde los League Indexes del 1, 2 y 3 representan la clase 1, los League Indexes 4, 5 y 6 representan la clase 2 y los los League Indexes 7 y 8 la clase 3. Se pide:

1. Diseñar un clasificador lineal basado en mínimos cuadrados para los dos problemas de clasificación.
2. Diseñar un discriminante lineal de Fisher, basado en el modelo de perceptrón o basado en SVM para el problema (i).
3. Diseñar un clasificador Bayesiano lineal para los dos problemas de clasificación. En este caso utilice a lo más dos características o variables independientes. Justifique adecuadamente las que usó.
4. *Optional*: diseñar un clasificador de Neyman-Pearson para para el problema (i). Justifique adecuadamente las que usó.
5. Diseñar un clasificador no supervisado basado en  $k$ -means usando la implementación de Matlab `kmeans`.

Se pide además comparar los desempeños de los clasificadores en términos de la probabilidad de clasificación y de error. Justifique además si se cumplen los supuestos impuestos por cada uno de los clasificadores.

La presentación de sus resultados debe hacerse en un informe de a lo más 5 páginas, donde la primera página debe ser un resumen ejecutivo (averiguar que se entiende por resumen ejecutivo) y las cuatro restantes en un formato a elección. Estas cuatro páginas deben contener al menos las siguientes secciones: (i) resumen, (ii) descripción muy breve del problema; (iii) solución(es) al problema, donde se debe incluir la selección de variables, tipo de regresión(es), tipo de función usada, etc. y se puede incluir, si es que sirve, diagramas de las variables, clústers, etc., (iv) resultados, y (v) conclusiones.