

# Práctico Aprendizaje Automático - v0.1

Aprendizaje Automático sobre Grandes Volúmenes de Datos

October 14, 2014

<http://aprendizajengrande.net/practico2.pdf>

## 1 Ejercicio CAP

De las tres capacidades del teorema CAP, ¿cuáles dos están presentes en el formalismo MapReduce? Discuta sistemas alternativos similares a MapReduce que tomen las otras dos combinaciones de capacidades. ¿Son posibles? ¿En qué situaciones prácticas serían útiles?

## 2 Ejercicio *Random Forest*

Repetir Ejercicio 2 del Práctico 1 usando una *random forest* de tres árboles sobre un 50% de los datos de entrenamiento y dos features al azar en cada nodo. Comparar precision/recall con respecto a sus resultados anteriores y al árbol óptimo. Usar crossvalidación en dos segmentos.

## 3 Ejercicio MapReduce

Implementar conceptualmente (pseudo-código) en el formalismo MapReduce los cálculos de *smoothing* usando cross-validación descritos en el paper *Good-Turing Smoothing without Tears* (Gale, 1995).

## 4 Ejercicio Descomposición LU

Para una matriz de 10,000x10,000, diagramar las ejecuciones del algoritmo de descomposición LU visto en clase, asumiendo los mismos parámetros de Xiang (2013). Indicar claramente qué tareas son map, reduce o ejecutadas en el nodo central y sobre qué parte de la matriz trabajan.

## 5 Ejercicio ActiveMQ

Implementar conceptualmente (pseudo-código) en el formalismo ActiveMQ un sistema MapReduce con localidad de datos y las mismas garantías de tolerancia a fallas que Hadoop.