

# TAREA 4 REDES NEURONALES AVANZADAS DE APRENDIZAJE PROFUNDO

Ing. Alejandro Ferreira Vergara  
Departamento de Ingeniería Matemática  
Universidad de La Frontera

24 de Agosto de 2020

Siempre utilizando la base de datos "sms-spam":

## Pregunta 1

Tokenice utilizando  $n$ -gramas, con  $n = \{2, 3\}$  y vectorice usando el método TF-IDF. Luego haga uso del algoritmo LDA (Análisis de Discriminante Lineal) para clasificar. Haga una comparación con los resultados obtenidos en clases.

## Pregunta 2

Vectorice utilizando el método TF-IDF y aplique PCA (Análisis de Componentes Principales), con  $num - componentes = \{8, 32\}$ . Luego haga uso del algoritmo LDA para clasificar. Haga una comparación con los resultados obtenidos en clases.

## Pregunta 3

Vectorice utilizando el método TF-IDF y aplique SVD (Descomposición en Valores Singulares) Truncado, con  $num - componentes = \{8, 32\}$ . Luego utilice LDA para clasificar. Haga una comparación con los resultados obtenidos en clases.

## Pregunta 4

Vectorice utilizando el método Bolsa de Palabras y aplique PCA, con  $num - componentes = \{8, 32\}$ . Luego utilice LDA para clasificar. Haga una com-

paración con los resultados obtenidos en clases.

## Pregunta 5

Vectorice utilizando el método Bolsa de Palabras y aplique SVD Truncado, con  $num - componentes = \{8, 32\}$ . Luego utilice LDA para clasificar. Haga una comparación con los resultados obtenidos en clases.

## Pregunta 6

Vectorice utilizando el método Bolsa de Palabras y aplique LDiA (Asignación Latente de Dirichlet), con  $num - componentes = \{8, 64\}$ . Luego utilice LDA para clasificar. Haga una comparación con los resultados obtenidos en clases.

## Pregunta 7

Tokenice utilizando  $n$ -gramas, con  $n = \{2, 3\}$ , vectorice utilizando el método Bolsa de Palabras y aplique LDiA (Asignación Latente de Dirichlet), con  $num - componentes = \{32\}$ . Luego utilice LDA para clasificar. Haga una comparación con los resultados obtenidos en clases.

INDICACIÓN: En cuanto a la comparación con los resultados de las clases, se deben comparar las dimensiones de la matriz  $X$  y los errores (Matriz de Confusión y/o Exactitud) de sus modelos de clasificación con los modelos expuestos en clases. Recuerde dividir su conjunto de entrenamiento en los conjuntos *Train* y *Test* o aplique la metodología Cross Validation.

Se asignará un problema al azar para cada estudiante, el que deberá ser resuelto y enviado como un archivo Jupyter Notebook a más tardar el día domingo 30 de agosto a las 18:00 hrs (al campus virtual) y presentado en clases el día lunes 31 de agosto.