

# Ciencia de Datos

## Tarea 7

Para entregar el 24 de mayo de 2019

1. Aunque hay varias extensiones de AdaBoost al caso multiclase, una de las mas usadas es la llamada SAMME (Stagewise Additive Modeling using a Multi-class Exponential loss function), ya que está basada en la caracterización estadística de Friedman et al. Implementa esta versión de AdaBoost y verifica su desempeño en un conjunto de datos con más de dos categorías.

Incluye una breve descripción del método basandote en el artículo: Zhu J, Zou H, Rosset S, and Hastie T (2009). *Multi-Class AdaBoost*. Statistics and Its Interface, 2, 349360. Puedes usar también los datos que ahí se muestran para reproducir los resultados.

2. Usando los datos de los dígitos escritos a mano y digitalizados (`mnist`), complementa el ejercicio que hiciste en la tarea 2 aplicando métodos de clasificación basados en
  - Redes neuronales
  - Máquinas de Soporte Vectorial
  - Árboles de clasificación
  - Random Forest
  - AdaBoost

Utiliza  $K$ -Fold CV como criterio para elegir el mejor modelo, así como para compararlos. ¿Qué método elegirías?

Especifica los parámetros que usaste en cada método de clasificación. Incluye gráficos *informativos* sobre el desempeño de cada método. Actualiza tu aplicación interactiva, si es que la implementaste en la primera tarea.

3. Repite el ejercicio 2 para los datos de frutas que usaste en la tarea 4. Utiliza la representación en el espacio HSV con la mediana y los cuartiles centrales.

**Puntos extra:** verifica el desempeño en del clasificador que elegiste en ejemplos “reales”. Toma algunas fotos de frutas y realiza un preprocesamiento básico para clasificarlas. Puedes usar el código en C (cortesía de Karen) para quitar el fondo de tu foto. Lee las instrucciones que vienen documentadas.

¿Cómo funciona tu clasificador? ¿Qué tan sensible es a las condiciones de la imagen (tamaño, rotación, etc.)?