

Tarea 2: Aprendizaje bayesiano

Objetivo

Este laboratorio tiene por objetivos: a) implementar el algoritmo Naive Bayes; b) aplicar herramientas de metodología; y c) analizar los resultados obtenidos.

Problema

Trabajaremos con el mismo conjunto de datos utilizado en la tarea 1: «AIDS Clinical Trials Group Study 175»¹

Se deberá:

- Implementar el algoritmo Naive Bayes visto en clase, agregando el siguiente hiperparámetro:

m : tamaño equivalente de muestra

Entrene y evalúe el modelo utilizando los siguientes cuatro valores del hiperparámetro m : 1, 10, 100 y 1000.

- Repase las herramientas de metodología vistas, discutiendo cuáles son útiles y cuáles no lo son para este problema; en particular aplicar técnicas de feature selection, cross validation, y presentar la matriz de confusión y la curva de precision-recall.
- Analice los resultados obtenidos, comparando con el resultado obtenido en la tarea anterior.

Se podrá utilizar pandas y scikit-learn para la carga del dataset, su preprocesamiento, división del conjunto de datos y para la evaluación (precision, recall, matriz de confusión, curva precision-recall).

Entregables

- Informe con las pruebas realizadas y los resultados obtenidos. El informe a entregar debe ser un Jupyter Notebook.
- Código escrito para resolver el problema, que debe poder ejecutarse sin dependencias locales.

Fecha límite de entrega

Miércoles 18 de setiembre (inclusive).

¹<https://archive.ics.uci.edu/dataset/890/aids+clinical+trials+group+study+175>