RESUMEN

[Tp N°5 2](#_Toc103873969)

[Ejercicio 5.4 2](#_Toc103873970)

[TP N°4 3](#_Toc103873971)

[Ejercicio 4.2 3](#_Toc103873972)

[Ejercicio 4.7 3](#_Toc103873973)

[Tp N°3 4](#_Toc103873974)

[Ejercicio 3.3: 4](#_Toc103873975)

[Tp N°2 5](#_Toc103873976)

[Ejercicio 1: En 5](#_Toc103873977)

[Ejercicio 2: 5](#_Toc103873978)

[Ejercicio 3 (optativo): 5](#_Toc103873979)

# Tp N°5

Elegir uno de entre los siguientes proyectos sugeridos y aplicar

clustering para resolverlo. Cada proyecto tiene datasets de diferentes características y dificultad (indicada con un semaforito).

Para los más sencillos trabajar individualmente y aplicar más de un método de clustering, utilizar diferentes parámetros, comparar evaluaciones. Para los más complejos, trabajar en grupos de dos personas y aplicar un único modelo.

Si encuentran algún otro caso de clústering (que no esté totalmente resuelto) lo pueden proponer para negociarlo con la cátedra.

## Ejercicio 5.4

Chatbot: El objetivo consiste en crear un chatbot utilizando patrones de entrada y respuestas predefinidos. <https://github.com/manlio99/Materia-de-aprendizaje/tree/master/3_MidtermProjects/ProjectPCB>

# TP N°4

Elegir uno de entre los siguientes proyectos (se agregan links a las NB correspondientes) y aplicar clasificación para resolverlo. Cada proyecto tiene datasets de diferentes características (texto, imágenes, audio, etc.) por lo que tiene diferente dificultad (indicada con un semaforito).

Para los más sencillos trabajar individualmente y aplicar más de un clasificador, utilizar diferentes parámetros, comparar evaluaciones.

Para los más complejos, trabajar en grupos de dos personas y aplicar un único modelo.

## Ejercicio 4.2

COVID-19 Detection in Chest X-Ray Images: El dataset tiene 600+ imágenes en tres carpetas (COVID-19, normal, neumonía). Desarrollar un extractor de características de las imágenes, y luego entrenar un clasificador.

<https://github.com/manlio99/Materia-de-aprendizaje/blob/master/3_MidtermProjects/ProjectCXR>

## Ejercicio 4.7

Señales camineras: El objetivo es obtener un modelo que determine el tipo de señal de tráfico, donde la dificultad es que las tomas fueron realizadas en condiciones realistas. Se provee un dataset etiquetado y dividido en las tres partes (train, validate, test).

<https://github.com/manlio99/Materia-de-aprendizaje/blob/master/3_MidtermProjects/ProjectRTS>

# Tp N°3

## Ejercicio 3.3:

En el notebook <https://github.com/manlio99/Materia-de-aprendizaje/blob/master/3_MidtermProjects/musica.ipynb>

hay un dataset con 2000+ canciones deSpotify de una usuaria, donde algunas fueron marcadas como gustadas y otras no. Cada canción tiene a su vez 16 atributos (nombre, artista, duración, bailable, etc.). Desarrollar un clasificador que prediga si una canción dada va a ser gustada o no. Aclarar y justificar los pasos, analizar y explicar los resultados.

1 usuario

16 variables

Mas de 20mil filas

[out: like, dislike] = Clasificador(in: tema)

Intro: En este ejercicio están dispuestas 16 variables a trabajar. En este universom hay que tratar de identificar los cluster like, dislike.

Puede haber variables que mejor clasifiquen

## Revisión Unidad 3

Parametrica: para la predicción se emplea una distribución de probabilidad conocida.

clasificadores no paramétricos

k-vecinos: no se entrena

definir metrica

Para cada candidato, se busca \_k\_ vecinos mas cercanos, y se clasifica por mayoria.

# Tp N°2

## Ejercicio 1: En

<https://github.com/manlio99/Materia-de-aprendizaje/blob/master/4_DataWrangling/data/casos_covid_bahia.csv>

y

<https://github.com/manlio99/Materia-de-aprendizaje/blob/master/4_DataWrangling/data/camas_covid_bahia.csv>

están datasets de la situación epidemiológica y sanitaria relacionada con el covid en Bahía Blanca en los últimos meses (fuente https://datos.bahia.gob.ar/dataset?groups=covid19).

Las variables son discretas (personas) pero se puede aplicar razonablemente el análisis estadístico como si fuesen valores continuos.

Hay variables que puedan ser razonablemente consideradas normales? Hay variables que tienen un comportamiento sospechoso?

Ejercicio 2:

Generar un dataset similar al de la pág. 12 de este apunte (dos conjuntos Gaussianos con diferente media y DS, N=50 c/u, uno con etiqueta A y otro con etiqueta B).

Utilizar la curva ROC para proponer un umbral para un clasificador por mínimo error. Evaluar algunos de los parámetros de calidad (exactitud, precisión, f-measure).

Recalcular para algunas variantes (por ejemplo, acercando las medias de los grupos A y B, cambiando el valor umbral, etc.).

Cómo podrían hacer no supervisado este proceso?

## Ejercicio 3 (optativo):

Cómo serían los pasos si el dataset fuese 2D?

Puedo utilizar la curva ROC sin ninguna estimación previa?

Caso negativo, qué estimación sería útil (y cómo obtenerla)?