

# Lineamientos para la evaluación del curso Inteligencia Artificial Supervisados

O. L. Quintero

26 de abril de 2023

## Resumen

Este documento contiene los lineamientos para la presentación de la actividad evaluativa.

## 1. Introducción

El objetivo del curso es proveer elementos teóricos y conceptuales que le permitan a los estudiantes de Inteligencia Artificial de Ingeniería Matemática, enfrentar el problema de construir un modelo compacto (learning machine) que permita representar fenómenos del mundo real.

Consecuentemente, los principios de teoría de aprendizaje fueron adelantados en la clase. Se debe cuestionar y NO desviar la tarea de aprendizaje automático, es decir no "presuponer" la naturaleza del mismo. Debe explorarse la construcción de diversos modelos mediante la aplicación de los conceptos.

El curso pretende adelantar la correcta aplicación de conceptos que permiten construir el modelo, evaluarlo y juzgar con perspectiva científica su desempeño.

Esta actividad evaluativa consiste en entrenar varios modelos de aprendizaje supervisado

1. Arbol de decision
2. Maquina de soporte vectorial con tres kernels: lineal, polinomico y base radial
3. Modelos de regresion lineal y/o logistica

Los conceptos generales se pueden revisar directamente del libro "Machine Intelligence for decision making" (en borrador para uso de los estudiantes de este curso y bajo edicion por Springer), y de las diversas fuentes citadas en el libro con artículos científicos y otros libros mas especializados en cada tema.

Para comenzar a realizar su proceso de aprendizaje (me refiero a practicar en datos juguete antes de abordar el problema real), el estudiante puede usar conjuntos de datos sintéticos IRIS DATASET.

## 2. Contenidos

El estudiante debe seleccionar un problema relevante, y realizar este ejercicio de manera INDIVIDUAL.

El estudiante puede usar cualquier tipo de programa para realizar el ejercicio que se detallará en la siguiente sección.

Los estudiantes abordaran el entrenamiento de modelos de aprendizaje siguiendo los siguientes elementos:

1. Error real deseado
2. Error de entrenamiento deseado
3. Garantía probable de aprendizaje
4. Tamaño óptimo de la muestra
5. Dividir el conjunto de muestra en: entrenamiento, validación y prueba.
6. Visualizar los datos con los algoritmos de visualización por defecto de la suite, extraer características si hace falta para aumentar la dimensionalidad
7. Visualizar los datos con el embebimiento BH tsne <https://lvdmaaten.github.io/tsne/> o con el UMAP que me gusta tanto
8. Entrenar el modelo con los datos en dimensiones originales
9. Entrenar el modelo con los datos en altas dimensiones (espacio original)
10. Entrenar el modelo con los datos en dimensiones reducidas (luego del embebimiento o con el autoencoder)
11. Realizar la comparación de los modelos en los espacios analizados

### 2.1. Entrenamiento de las maquinas de aprendizaje

Para su problema con  $m$  entradas,  $n$  salidas y  $N$  cantidad de instancias (patrones o tamaño de la muestra). Seleccionar las siguientes maquinas:

1. Regresión lineal o regresión logística
2. Árbol de decisión
3. Máquina de soporte vectorial con kernel lineal
4. Máquina de soporte vectorial con kernel polinómico
5. Máquina de soporte vectorial con kernel base radial

Plantear el problema usando las garantías para estas máquinas y su dimensión VC. En este dibujito la tarea que puse en el 2022-2, pero pueden guiarse por el tablero de la clase que hicimos.

# Trabajo Supervisado 28 de agosto

- Hacerlo en formato IEEE **Objetivo:** Desarrollar un conjunto de modelos de aprendizaje supervisado mediante la evaluación de conceptos relacionados con las PA6 y la relación de la complejidad del espacio generado y  $M, H, n, \epsilon, \delta$ .

## Partes

① **Introducción:** Contexto del problema, ¿Qué es lo último que se hizo con este problema en aprendizaje supervisado?

② **Marco conceptual:** Escribir, construir una definición de  $\epsilon, \delta, H, PA6, n, \Omega, (\Omega, \epsilon, \sigma), |S|$

## ③ Método

3.1 **Muestreo** Copiar pseudocódigo del proceso de muestreo con las 3 distribuciones

3.2 **Visualización**

3.3 **Disposición final de |S|** en qué estado quedaron mis clases, mis datos.

3.4 **Consideraciones de Aprendizaje** decir mis parejas de error  $(\epsilon, \delta)$  (son 3) y por qué las escogí en términos del problema.

También definir las PA6 para las distintas máquinas

• Árbol;

• SVM { Kernel: • lineal • Base Radial • Polinomio de cualquier grado

• Regresión lineal

• Regresión logística

Todas estas máquinas ya tienen definida la dim VC

Adicional, trabajamos con la Máquina propia, la creada : ( y creamos la siguiente tabla.

	$M_1$	$M_2$	$M_3$	$M_4$	$M_5$	$M_6$	$M_7$
$\epsilon_1 \delta_1$	1	1	1	1	1	1	1
$\epsilon_2 \delta_2$							
$\epsilon_3 \delta_3$							

Puede ocurrir en un caso que no me alcancen los datos, solo hago los casos en los que sí me alcance.

## ④ Resultados

4.1 **Medir desempeño** Solo con los errores  $\delta$  y  $\epsilon$

$\delta$  Calculo  $|S \cap (h \Delta G)|$   
|S|

4.2 (Supongo que es del 4.1) Calculo  $\epsilon P(h \Delta G_i)$  con el 20% de datos TEST, si con este me da mal, lo hago con el 20% de validación, si todavía no me da, remuestro los datos y si nada abordo la situación en la próxima parte 4.3

## 4.3 Análisis

## ⑤ Conclusiones

En Términos del código, en la parte inicial poner todos los requirements del código, que se le instalen a Olga :)  
Poner un path general para los datos local o archivos para nube :)

Figura 1: Tablerito

## 2.2. Entregables

Se debe entregar:

- Documento informe
- Codigos elaborados

Por favor asistir a asesoría con el profesor del curso en la plataforma teams y solicitar su cita con anticipación al correo oquintel@eafit.edu.co.