

Laboratorio de aprendizaje de máquinas

Práctica 1: Clasificador Perceptrón Universidad Tecnológica de Pereira, Facultad de ingenierías, Maestría en Ingeniería de Sistema

Msc Ing Carlos Alberto Henao Baena
caralbhenao@utp.edu.co

Objetivo

Al finalizar esta práctica el estudiante conocerá e identificará los elementos básicos en una tarea de clasificación lineal en un paradigma biclase y multiclase aplicando el algoritmo Perceptrón, también evaluará el desempeño mediante métricas de rendimiento para este tipo de tarea.

Laboratorio (Clasificador perceptrón)

Considere el dataset *DataLeastSquareClasificationTwoClass.csv* y desarrolle los siguientes pasos en un *Google Colab Notebook*.

1. Transfiera el dataset a un frame panda utilizando el método `pd.read.csv('dirección_dataset_cargado.csv')`
2. Divida el dataset de la siguiente forma: 75% para entrenamiento el restante para validación
3. Inicialice el vector de parámetros W utilizando algún método de muestreo estadístico, recuerde que para un problema biclase W este tiene tres valores.
4. Establezca un valor de *LearningRate* = 0.3
5. Implemente el algoritmo de perceptron para actualizar los valores del vector de parámetros W (puede utilizar cualquiera de las alternativas estudiadas en clase)

Una vez el algoritmo de perceptrón finalice responder desarrollar lo siguiente:

6. Grafique en un mismo gráfico lo siguiente:
 - $X1_{Train}$ vs $X2_{Train}$
 - $X1_{Test}$ vs $X2_{Test}$
 - El plano de decisión y_0 que para un caso lineal viene dado por $y_0 = -\frac{W[2]+W[0]X1_{Train}}{W[1]}$
7. Realice la predicción sobre el conjunto de datos de entrenamiento y validación utilizando la ecuación 2.
8. Realice la predicción sobre el conjunto de datos de entrenamiento y validación utilizando la ecuación 1.

$$\hat{H} = f(W^T X) = \begin{cases} 1 & W^T X > 0 \\ -1 & W^T X \leq 0 \end{cases} \quad (1)$$

9. Conocidas las predicciones sobre el conjunto de datos de Test evalúe las siguientes métricas:

- Matriz de confusión
- La sensibilidad
- Especificidad
- Precisión
- Exactitud

10. Repita el procedimiento para la base de datos *DataLeastSquareClasificationTwoClassOutliers.csv*
11. Implemente un clasificador multiclase (realizando los cambios respectivos en el algoritmo de aprendizaje) utilizando la base de datos *DataClasificationMultiClass.csv*

Realice un informe técnico documentado la experiencia (introducción, metodología, resultados, conclusiones y bibliografía)