

Práctica 2

Perceptrón



Objetivos

El objetivo principal de esta práctica es comprender el funcionamiento del perceptrón como bloque de construcción elemental de las redes neuronales. Adicionalmente se realiza un repaso sobre el análisis y preprocesamiento de los datos.

Temas

- Perceptrón. Entrenamiento.
- Normalización y tratamiento de datos
- Correlación
- Representaciones Gráficas

Lectura

Material de Lectura: Capítulo 13 del Libro Introducción a la Minería de Datos de Hernández Orallo

Ejercicio 1

La Tabla 1 muestra información correspondiente a la resistencia adquirida por piezas de cierto material luego de haber sido sometidas a distintas temperaturas

TEMPERATURA	DUREZA	RESISTENCIA
-12	ALTA	SUPERIOR
20	MEDIA	SUPERIOR
0	ALTA	SUPERIOR
7	MEDIA	SUPERIOR
11	BAJA	NORMAL
-30	ALTA	SUPERIOR
63	BAJA	NORMAL
94	BAJA	NORMAL
45	MEDIA	NORMAL

Tabla 1

Se numerizó el atributo DUREZA de la siguiente forma: BAJA \rightarrow 0, MEDIA \rightarrow 30, ALTA \rightarrow 60

- Luego de la numerización se calculó el coeficiente de correlación lineal entre los atributos TEMPERATURA y DUREZA y se obtuvo como resultado -0.78. ¿Cómo debe interpretarse este valor?
- Luego de numerizar el atributo DUREZA, los ejemplos fueron utilizados para entrenar un perceptrón capaz de predecir correctamente el atributo RESISTENCIA. Los pesos obtenidos fueron los siguientes:

$$w(\text{TEMPERATURA}) = 2.386; w(\text{DUREZA}) = -2.196; b = -0.023$$

¿Cuál será la respuesta del perceptrón para un material sometido a una TEMPERATURA de 9 grados Celsius que presenta una DUREZA =BAJA?

Ejercicio 2

Se desea construir una Red Neuronal, formada por un único Perceptrón, para clasificar fotos de hojas diferenciando las que corresponden a helechos de las que no. A continuación, se muestran algunas de las imágenes que se utilizarán en el entrenamiento:



Se analizaron las imágenes y se extrajeron características geométricas representativas. El archivo **hojas.csv** contiene la cantidad de píxeles correspondientes al perímetro y el área de cada hoja.

- Utilice los ejemplos del archivo **hojas.csv** para entrenar un perceptrón que permita reconocer cuando se trata de una hoja de helecho. Utilice una velocidad de aprendizaje (parámetro alfa) de 0.01 y una máxima cantidad de iteraciones MAX_ITE=300.
- A partir de los pesos del perceptrón entrenado, indique cuál es la función discriminante obtenida.

- c) Calcule manualmente la respuesta del perceptrón si se ingresa una hoja con un perímetro de 770 píxeles y un área de 5000 píxeles.
- d) Realice 50 ejecuciones independientes del entrenamiento de a) utilizando una máxima cantidad de iteraciones $MAX_ITE=100$. Luego complete el siguiente cuadro considerando sólo los casos exitosos en los que se logró obtener un perceptrón capaz de clasificar correctamente todos los ejemplos. Registre el porcentaje de ejecuciones con un accuracy del 100% y la cantidad de iteraciones promedio empleadas en estas ejecuciones exitosas.

Normalización de los datos de entrenamiento	Orden de los datos en el entrenamiento	Velocidad de aprendizaje	% Ejecuciones en las que se alcanzó el 100% de acierto	Promedio de iteraciones en las ejecuciones que alcanzaron el 100%
Sin normalizar	Aleatorio	0.2		
		0.005		
	Ascendente	0.2		
		0.005		
	Descendente	0.2		
		0.005		
Lineal Uniforme	Aleatorio	0.2		
		0.005		
	Ascendente	0.2		
		0.005		
	Descendente	0.2		
		0.005		
Usando media y desvío	Aleatorio	0.2		
		0.005		
	Ascendente	0.2		
		0.005		
	Descendente	0.2		
		0.005		

- e) En base a los resultados del cuadro anterior
- Analice la importancia de normalizar los datos de entrada.
 - La relación entre la velocidad de aprendizaje y la cantidad de iteraciones.
 - Compare los resultados obtenidos con las dos normalizaciones.
 - ¿Considera que los resultados del algoritmo se ven afectados por el orden en el que se ingresan los ejemplos?

Ejercicio 3

El archivo **SEMILLAS.csv** contiene información de granos que pertenecen a tres variedades diferentes de trigo: Kama, Rosa y Canadiense. El total es de 210 ejemplos a razón de 70 ejemplos para cada tipo de grano, seleccionados al azar para el experimento. La información registrada corresponde al resultado de la visualización de alta calidad de la estructura interna del núcleo efectuada utilizando una técnica de rayos X blandos. Este tipo de estudio no es destructivo y es considerablemente más económico que otras técnicas de imagen más sofisticadas como la microscopía de barrido o la tecnología láser. Las imágenes se grabaron en placas KODAK de rayos x de 13x18 cm. Los estudios se realizaron utilizando granos de trigo cosechados combinados procedentes de campos experimentales, explorados en el Instituto de Agrofísica de la Academia Polaca de Ciencias en Lublin. Para construir los datos, se midieron siete parámetros geométricos de cada grano de trigo:

- i. área A,
- ii. perímetro P,
- iii. compacidad $C = 4 * \pi * A / P^2$,
- iv. longitud del núcleo,
- v. ancho del núcleo,
- vi. coeficiente de asimetría
- vii. longitud del surco del núcleo

A partir de los 210 ejemplos, luego de normalizarlos utilizando los valores de media y desvío, se logró entrenar un perceptrón capaz de identificar, con una precisión del 100%, uno de los tres tipos de semillas. Para realizar el entrenamiento se utilizó una velocidad de aprendizaje de 0.05 y un máximo de 200 iteraciones. Indique cuál es el tipo de semilla que puede ser reconocido correctamente por un perceptrón.

Ejercicio 4

El archivo **ZOO.csv** contiene información de 101 animales caracterizados por los siguientes atributos

A1. Nombre del animal	A7. Acuático	A13. Aletas
A2. Tiene Pelo	A8. Depredador	A14. Patas
A3. Plumas	A9. Dentado	A15. Cola
A4. Huevos	A10. Vertebrado	A16. Domestico
A5. Leche	A11. Branquias	A17. Tamaño gato
A6. Vuela	A12. Venenoso	A18. Clase

Salvo los atributos A1 y A18 que contienen texto y el A14 que contiene el número de patas del animal, el resto toma el valor 1 si el animal posee la característica y 0 si no. Hay 7 valores de clase posible (atributo A18): mamífero, ave, pez, invertebrado, insecto, reptil y anfibio.

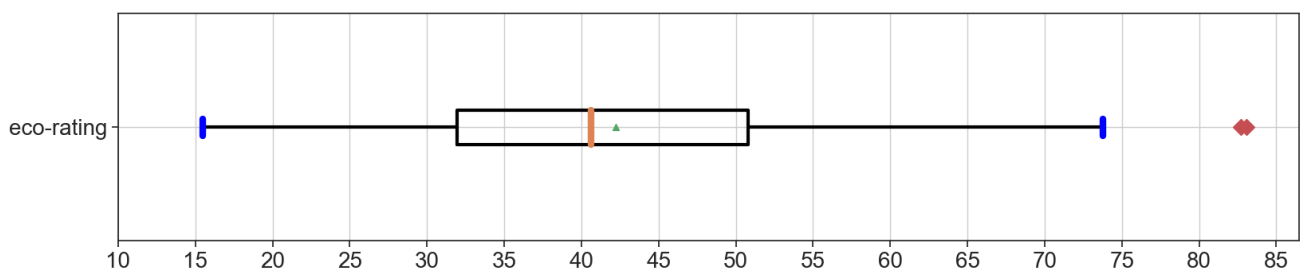
- a) Realice un gráfico que visualice de la cantidad de ejemplos por cada valor del atributo **clase** y analice que tipos de problema podrían surgir al entrenar un modelo para clasificación.
- b) Utilice todos los ejemplos para entrenar un perceptrón que sea capaz de reconocer si un animal es un mamífero. Entrene varias veces si es necesario y verifique que funciona correctamente.
- c) Observe los pesos del perceptrón entrenado en a) ¿Puede determinar cuáles son las características más utilizadas para decidir si se trata de un mamífero o no? Realice varias ejecuciones independientes y observe si las características más utilizadas siguen siendo las mismas.
- d) Repita b) y c) para las aves.
- e) Repita b) y c) para los reptiles.

Ejercicio 5

El archivo **automobile-simple.csv** contiene 11 atributos de automóviles de un total de 205 registros. Es una versión modificada y simplificada del dataset disponible en el repositorio UCI <https://archive.ics.uci.edu/ml/datasets/Automobile>. La siguiente tabla contiene una breve descripción de los atributos que contiene el archivo y caracterizan a cada vehículo.

Atributo	Descripción
make	marca: alfa-romero, audi, bmw, chevrolet, dodge, honda, isuzu, jaguar, mazda, mercedes-benz, mercury, mitsubishi, nissan, peugot, plymouth, porsche, renault, saab, subaru, toyota, volkswagen, volvo
fuel-type	tipo de combustible: diesel, gasolina.
num-of-doors	número de puertas: cuatro, dos.
body-style	tipo de carrocería: techo duro, wagon, sedán, hatchback, descapotable.
curb-weight	peso en vacío: numérico de 1488 a 4066.
engine-size	tamaño del motor: numérico de 61 a 326.
horsepower	potencia: numérico de 48 a 288.
peak-rpm	velocidad máxima: numérico de 4.150 a 6.600 rpm.
city-mpg	Rendimiento en ciudad (en millas por galón): numérico de 13 a 49.
highway-mpg	Rendimiento en ruta (en millas por galón): numérico de 16 a 54.
price	Precio en USD: numérico de 5118 a 45400.
volume	Volumen del vehículo (alto x ancho x alto).
eco-rating	Valoración de que tan ecológico es el vehículo en función del volumen, peso, consumo y tipo de combustible.

- Para cada atributo indique si es Discreto, Continuo, Nominal u Ordinal.
- Analice y Prepare los atributos del dataset:
 - Calcule la correlación de los atributos y complete los valores faltantes. Basándose en la matriz de correlación, asigne valores representativos para **price** y **horsepower**. Compare dichos valores con la media y la mediana de estos atributos y reflexione sobre su importancia.
 - Dado el siguiente diagrama de caja (boxplot) del atributo eco-rating, complete la siguiente tabla aproximando los valores visualmente y luego complete con los valores correctos



Medida	Valor/es aproximado/s	Valor/es correcto/s
Mínimo		
Máximo		
Q1		
Q2 o mediana		
Q3		
RIC		
Bigote superior		
Bigote inferior		
Intervalos de valores atípicos leves		
Intervalos de valores atípicos extremos		

- c) Realice el entrenamiento de un perceptrón para que aprenda a clasificar si un auto es ecológico. Tenga en cuenta los siguientes pasos:
- I. Utilice el atributo eco-rating para generar un nuevo atributo binario que determine si un auto es ecológico o no. Un auto es considerado ecológico si el valor de eco-rating supera la media de dicho atributo.
 - II. Genere y compare 3 modelos utilizando diferentes normalizaciones (Sin normalizar, normalización lineal, normalización estándar).
 - III. Teniendo en cuenta la matriz de correlación del punto b) repita el punto II) eliminando dos atributos fuertemente correlacionados (uno negativo y otro positivo). Compare y reflexiones sobre los resultados obtenidos