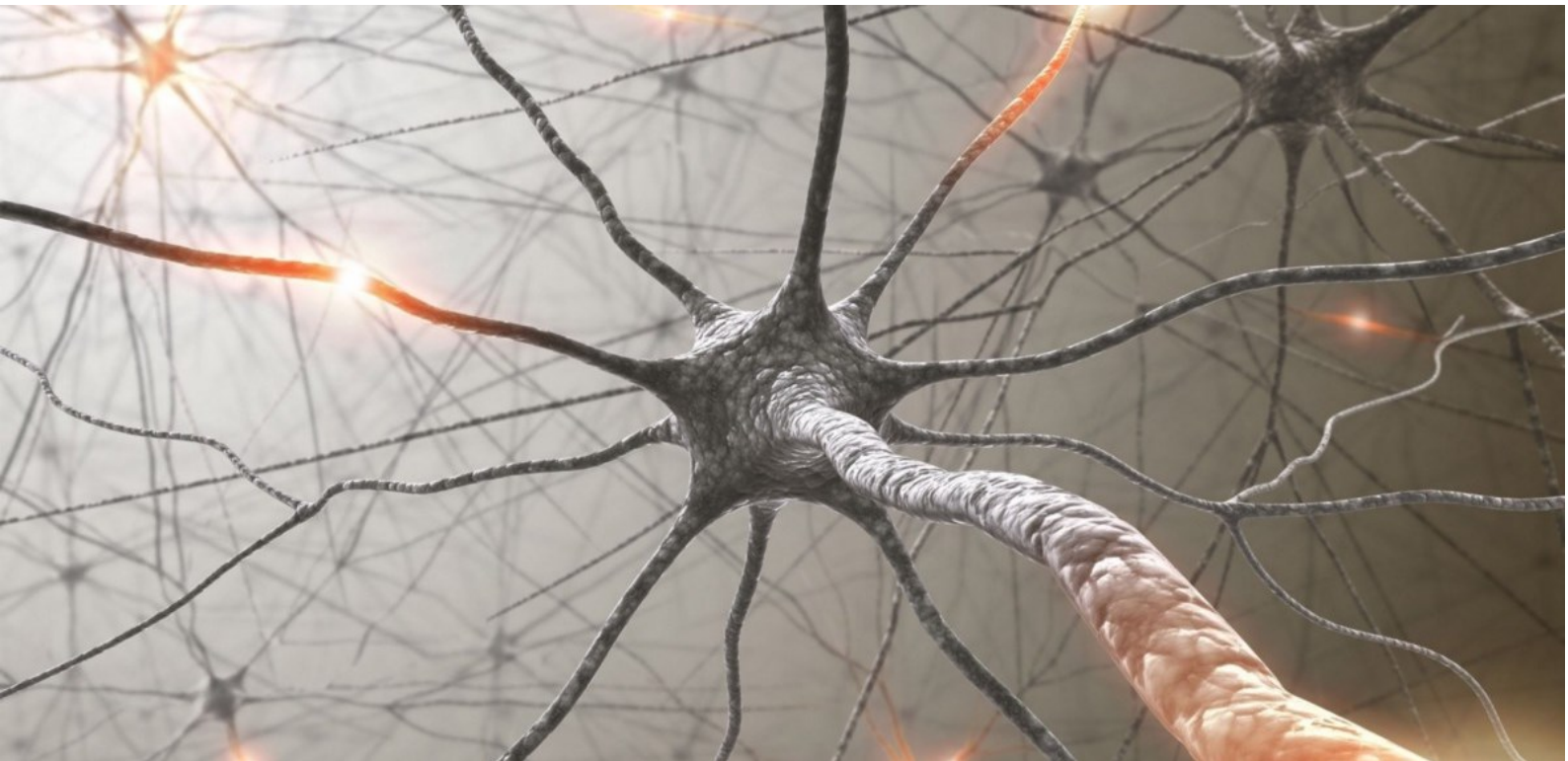


Práctica 2

Perceptrón



Objetivos

El objetivo de esta práctica es comprender el funcionamiento del perceptrón como bloque de construcción elemental de las redes neuronales.

Temas

- Perceptrón. Entrenamiento.
- Normalización de datos

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 distribución de cierto producto de una empresa

UNIDADES	DESCUENTO	ENVÍO
20	BAJO	NORMAL
29	MEDIO	NORMAL
45	BAJO	SUPERIOR
30	ALTO	SUPERIOR
30	MEDIO	NORMAL
2	BAJO	NORMAL
90	ALTO	SUPERIOR
120	ALTO	SUPERIOR
46	MEDIO	SUPERIOR

Tabla 1

Donde:

- **UNIDADES** es un atributo numérico que indica la cantidad de unidades del producto que deben ser entregadas.
- **DESCUENTO** es el nivel de descuento que recibirá el comprador en esta operación.
- **ENVIO** es el tipo de envío que el comprador eligió

Se numerizó el atributo DESCUENTO de la siguiente forma: BAJO → 0, MEDIO → 25, ALTO → 50

- a) Luego de la numerización se calculó el coeficiente de correlación lineal entre los atributos UNIDADES y DESCUENTO y se obtuvo como resultado 0.679. ¿Cómo debe interpretarse este valor?
- b) Luego de numerizar el atributo DESCUENTO, los ejemplos fueron normalizados linealmente en el intervalo [0; 1] y utilizados para entrenar un perceptrón capaz de predecir correctamente el atributo ENVIO. Los pesos obtenidos fueron los siguientes:

$$w(\text{UNIDADES}) = 0.484; w(\text{DESCUENTO}) = 0.06; b = -0.161$$

¿Cuál será la respuesta del perceptrón para un pedido de 41 UNIDADES con un DESCUENTO = BAJO?

- c) Luego de numerizar el atributo DESCUENTO de la siguiente forma BAJO → 0, MEDIO → 25, ALTO → 50 y de normalizar los ejemplos de la Tabla 1 utilizando los valores de media y desvío (tipificación) se entrenaron varios perceptrones con el objetivo de predecir el valor del atributo ENVÍO, pero no todos funcionaron correctamente. Indique cuáles de las siguientes funciones discriminantes corresponden a perceptrones que clasifican correctamente el valor del atributo ENVIO para los 9 ejemplos de la Tabla 1.
 - i. $0.043 * \text{UNIDADES} + 0.006 * \text{DESCUENTO} + 0.015 = 0$
 - ii. $\text{DESCUENTO} = -7.007 * \text{UNIDADES} - 2.345$
 - iii. $0.139 * \text{UNIDADES} + 0.018 * \text{DESCUENTO} + 0.030$
 - iv. $\text{DESCUENTO} = -1.816 * \text{UNIDADES} - 0.484$

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:



Las imágenes ya han sido analizadas. El archivo **hojas.csv** contiene la cantidad de pixeles 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.
- Calcule manualmente la respuesta del perceptrón si se ingresa una hoja con un perímetro de 770 pixeles y un área de 5000 pixeles.
- 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	34%	66 %
		0.005	20 %	80 %
	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		

- 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	A10. Vertebrado
A2. Tiene Pelo	A11. Branquias
A3. Plumas	A12. Venenoso
A4. Huevos	A13. Aletas
A5. Leche	A14. Patas
A6. Vuela	A15. Cola
A7. Acuático	A16. Domestico
A8. Depredador	A17. Tamaño gato
A9. Dentado	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) Utilice todos los ejemplos para entrenar un perceptrón que sea capaz de reconocer si un animal es un mamífero. Verifique que funciona correctamente.
- b) 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.
- c) Repita a) y b) para las aves.

Ejercicio 5

El archivo **Iris.csv** contiene información referida a la longitud y al ancho de sépalos y pétalos de tres especies de flores: *iris setosa*, *iris versicolor* e *iris virginica*.

Entrene un perceptrón que sea capaz de determinar si una flor pertenece a la especie *iris setosa* o no. Utilice el 70% de los ejemplos para entrenar y el 30% para realizar el testeo.

Realice al menos 10 ejecuciones independientes de la configuración seleccionada para respaldar sus afirmaciones referidas a la precisión del perceptrón tanto para el conjunto de entrenamiento como para el de testeo.

Ejercicio 6

El archivo **Sonar.csv** contiene registros de rebotes de señales de sonar tomadas en varios ángulos y bajo distintas condiciones.

La tarea es utilizar un perceptrón para discriminar entre señales de sonar rebotadas en un cilindro de metal y aquellas rebotadas en una roca más o menos cilíndrica.

Cada muestra es un conjunto de 60 números en el rango de 0 a 1. Cada número representa la energía dentro de una banda de frecuencia particular, integrada durante un cierto período de tiempo. La etiqueta asociada a cada registro contiene "Rock" si el objeto es una roca y "Mine" si es una mina (cilindro de metal).

Para ello utilice una parte de los registros para entrenar el perceptrón y el resto para realizar el testeo. Pruebe con distintos porcentajes e indique la precisión obtenida en la clasificación.

Realice al menos 10 ejecuciones independientes de c/u de las configuraciones seleccionadas para respaldar sus afirmaciones referidas a la precisión del perceptrón tanto para el conjunto de entrenamiento como para el de testeo.