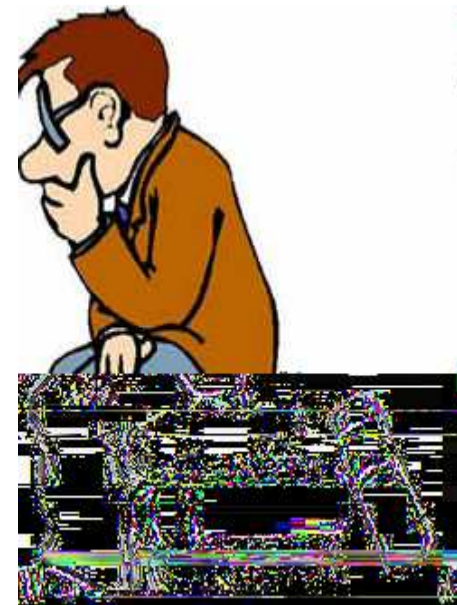


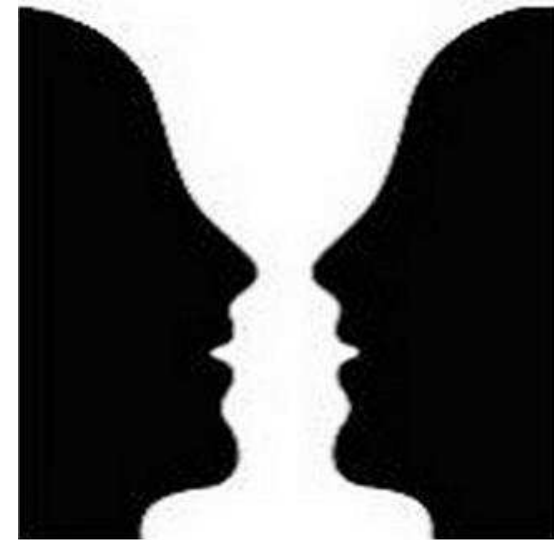
Aprendizaje de Maquinas

Jorge Luis Guevara Díaz

- ¿Es posible que las máquinas puedan asociar y aprender de la experiencia?



● ● ● | Como Asociamos?





imagine

- Computadoras que aprendan de registros médicos y brinden tratamiento mas efectivo a nuevas enfermedades
- Casas que automáticamente ahorren energía, basado en los patrones de uso de sus habitantes



- Lamentablemente todavia los científicos en computacion no logran hacer que las máquinas aprendan como lo hacen las personas
- Sin embargo existen buenos algoritmos para determinadas tareas de aprendizaje



Introduccion

- Es el campo que estudia los algoritmos y técnicas para permitir que las máquinas puedan aprender de la experiencia, y no solamente eso si no que los programas puedan mejorar con la experiencia.
- Existen muchos algoritmos, como las redes bayesianas, las maquinas de soporte vectorial, las redes neuronales, Modelos Ocultos de Markov etc,



Introduccion

- Definicion

- Un programa de computadora se dice que ***aprende*** de la experiencia ***E*** con respecto a una clase de tarea ***T*** y medida de performance ***P***, si su performance en la tarea ***T*** medida con ***P*** mejora con la experiencia ***E***



Introduccion

- Ciertos problemas no pueden ser solucionados por enfoques tradicionales en computación, como por ejemplo:

No se conoce en forma determinada un algoritmo para calcular la salidas de un conjunto de datos de entrada

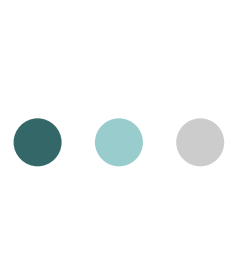
Calcular la salida deseada es muy costosa computacionalmente



Introduccion

○ Aplicaciones

- Clasificación de proteínas en base a su secuencia de ADN
- Modelar reacciones químicas donde las interacciones precisas de sus reactantes es desconocida
- Eliminación spam
- Biometría computacional
- Reconocimiento de patrones
- Minería de Datos



Ejemplos de programas exitosos

- Reconocimiento automatico del habla (Waibel 1989; Lee 1989)
- Prediccion de tasas de recuperacion de pacientes de neumonia (Cooper et al. 1997)
- Detectar uso fraudulento de tarjetas de crédito
- Manejo de vehiculos en forma autónoma en autopistas públicas (Pomerleau 1989)
- Juegos en el nivel competitivo de campeones humanos (Tesauro 1992, 1995)



Algunos términos

- Datos de entrenamiento
 - Muestras de entrada/salida en la fase de entrenamiento
- Metodología de aprendizaje
 - Enfoque utilizado para construir un maquina de aprendizaje de acuerdo a los datos



Algunos términos

- Función objetivo
 - Encontrar la función objetivo es la solución del problema de aprendizaje
- Hipótesis
 - Conjunto de funciones candidatas de las cuales se tratará de encontrar la función correcta
- Espacio de hipótesis
 - Conjunto de hipótesis



Algunos términos

- Algoritmo de aprendizaje
 - Es el algoritmo que toma de los datos de entrenamiento como entrada y selecciona una hipótesis del espacio de hipótesis
- Clasificación binaria
 - Problema de aprendizaje con salidas binarias, ejemplo clasificar rostros y no rostros salida si/no



Algunos términos

- Clasificación multiclase
 - Reconocer diferentes tipos de proteínas dada una secuencia de ADN
- Regresión
 - Salida dada en valores reales, por ejemplo encontrar los valores de salida que modelan una reacción química



Tipos de aprendizaje

- Supervisado

Se necesitan dar ejemplos previos de pertenencia a determinada clase de los valores de entrada

- No supervisado

los datos son incluidos en una clase en particular conforme avanza el proceso de aprendizaje



Generalizacion

- Es la habilidad de una hipótesis de clasificar correctamente datos que no pertenecen al conjunto de datos de entrenamiento
- Se desea optimizar esta función



Maquinas de aprendizaje lineales

- ***Clasificación lineal***

Problema:

para una entrada $\mathbf{X} = \{\mathbf{x}_1 \dots \mathbf{x}_n\}$

encontrar una funcion f (lineal)

tal que

$$f(\mathbf{x}) = 1 \quad \text{ssi } f(\mathbf{x}) \geq 0$$

$$f(\mathbf{x}) = -1 \quad \text{ssi } f(\mathbf{x}) < 0$$



Maquinas de aprendizaje lineales

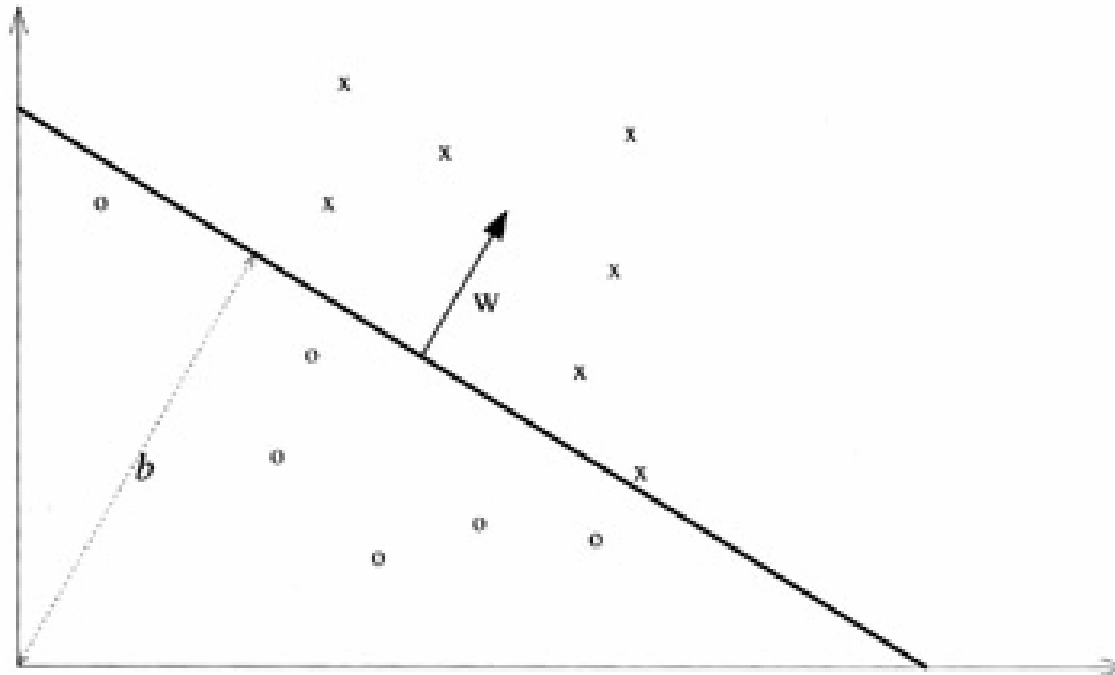
- w y b son parametros que controlan la funcion y pertenecen a los reales

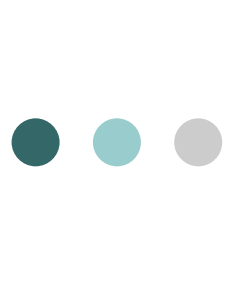
$$\begin{aligned} f(\mathbf{x}) &= \langle \mathbf{w} \cdot \mathbf{x} \rangle + b \\ &= \sum_{i=1}^n w_i x_i + b \end{aligned}$$



Maquinas de aprendizaje lineales

- Interpretacion geométrica de este espacio de hipótesis, para el espacio de entrada X





Maquinas de aprendizaje lineales

- La ecuacion del hiperplano de dimensión ***n-1*** es **$\langle w.x \rangle + b = 0$** , el cual divide el espacio de entrada en dos diferentes clases
- Las maquinas de aprendizaje lineales son las mas simples de estudiar, y son conocidas en la estadística como ***discriminantes lineales*** y en el campo de las redes neuronales como ***perceptrones***



Redes Neuronales

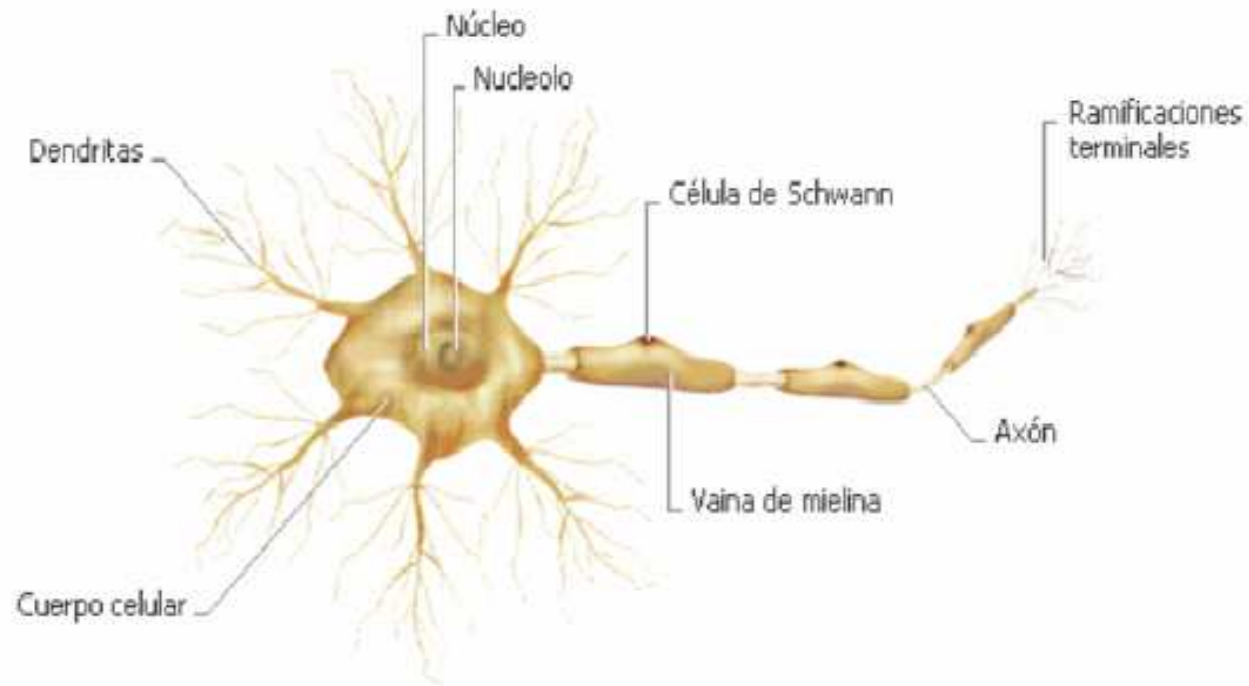
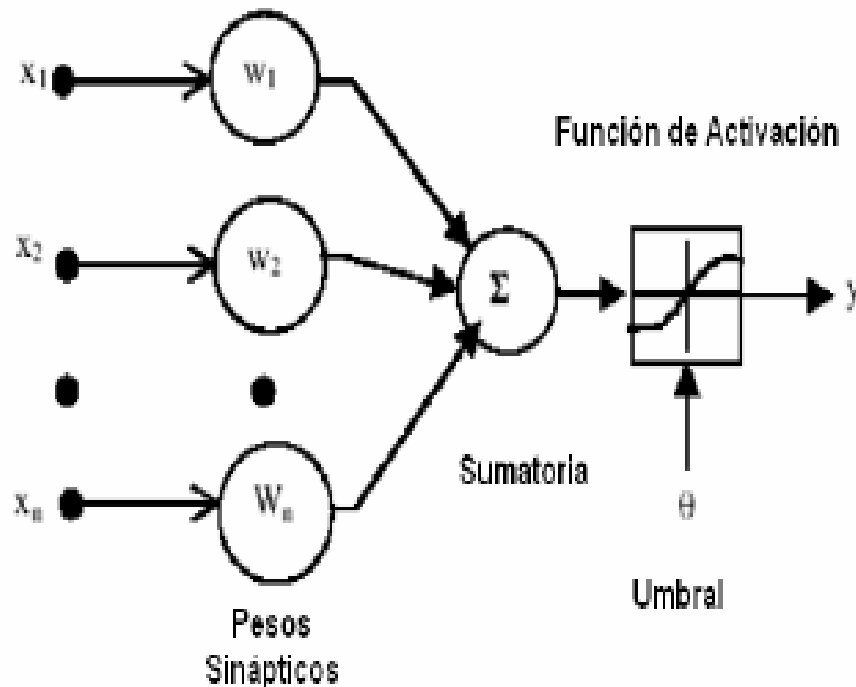


Figura 1: Una red neuronal



Redes Neuronales Artificiales

$$\text{Entrada Neurona}_i = \sum_{j=1}^n x_j w_{ij}$$



$$f(u) == \begin{cases} 0 & \text{si } u < 0 \\ 1 & \text{si } \geq 0 \end{cases}$$

$$f(u) == \begin{cases} -1 & \text{si } u < -1 \\ u & -1 \leq u \leq 1 \\ 1 & \text{si } \geq 1 \end{cases}$$

$$f_u = \frac{1}{1 + e^{-au}} \cdot \quad 0 \leq f(u) \leq 1$$

$$f_u = \frac{e^{\alpha u} - e^{-\alpha u}}{e^{\alpha u} + e^{-\alpha u}}$$



El perceptrón de Rosemblat

- El primer algoritmo para aprendizaje de clasificaciones lineales fue propuesto por Rosemblat en 1956.
- Dado un conjunto de entrenamiento

$$S = ((\mathbf{x}_1, y_1), \dots, (\mathbf{x}_\ell, y_\ell)) \subseteq (X \times Y)^\ell$$

El perceptrón de Rosemblat forma primal

- Dado un conjunto linealmente separable S y una tasa de aprendizaje $\eta \in \mathbb{R}^+$

$$\mathbf{w}_0 \leftarrow \mathbf{0}; b_0 \leftarrow 0; k \leftarrow 0$$

$$R \leftarrow \max_{1 \leq i \leq \ell} \|\mathbf{x}_i\|$$

repeat

for $i = 1$ to ℓ

if $y_i(\langle \mathbf{w}_k \cdot \mathbf{x}_i \rangle + b_k) \leq 0$ then

$$\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k + \eta y_i \mathbf{x}_i$$

$$b_{k+1} \leftarrow b_k + \eta y_i R^2$$

$$k \leftarrow k + 1$$

end if

end for

Hasta que no haya ningun error dentro del for

return (\mathbf{w}_k, b_k)

donde k es el número de errores



El perceptrón de Rosenblatt forma Dual

- Como se podrá observar el vector de pesos \mathbf{w} está expresado como combinación lineal de \mathbf{x} y de \mathbf{y}
- Una representación alternativa del perceptrón se tiene teniendo en cuenta lo siguiente:

$$\mathbf{w} = \sum_{i=1}^{\ell} \alpha_i y_i \mathbf{x}_i$$



El perceptrón de Rosenblatt forma Dual

- La función de decisión puede ser escrita en coordenadas duales

$$\begin{aligned}h(\mathbf{x}) &= \text{sgn}(\langle \mathbf{w} \cdot \mathbf{x} \rangle + b) \\&= \text{sgn}\left(\left\langle \sum_{j=1}^{\ell} \alpha_j y_j \mathbf{x}_j \cdot \mathbf{x} \right\rangle + b\right) \\&= \text{sgn}\left(\sum_{j=1}^{\ell} \alpha_j y_j \langle \mathbf{x}_j \cdot \mathbf{x} \rangle + b\right),\end{aligned}$$

- ● ● | El perceptrón de Rosemblat
forma Dual
 - Dado un conjunto de entrenamiento S

$\alpha \leftarrow \mathbf{0}; b \leftarrow 0$

$R \leftarrow \max_{1 \leq i \leq \ell} \|\mathbf{x}_i\|$

repeat

for $i = 1$ to ℓ

if $y_i \left(\sum_{j=1}^{\ell} \alpha_j y_j \langle \mathbf{x}_j, \mathbf{x}_i \rangle + b \right) \leq 0$ then

$\alpha_i \leftarrow \alpha_i + 1$

$b \leftarrow b + y_i R^2$

end if

end for

Hasta que no haya ningun error dentro del for

return (α, b) que define la función $h(\mathbf{x})$



El perceptrón de Rosemblat forma Dual

- Esta representación alternativa del perceptron con su función de decisión tiene propiedades interesantes, por ejemplo los puntos que son más difíciles de aprender tienen un valor alto de alfa



Convergencia del Perceptrón

- Prueba de Novikoff

- Sea \mathbf{S} un conjunto de entrenamiento no trivial

$$R = \max_{1 \leq i \leq \ell} \|\mathbf{x}_i\|.$$

- Suponer que existe un vector \mathbf{w}_{opt} : $\|\mathbf{w}_{opt}\| = 1$
- Además $y_i(\langle \mathbf{w}_{opt} \cdot \mathbf{x}_i \rangle + b_{opt}) \geq \gamma$
- El número de mistakes en el algoritmo del perceptrón es a lo mucho $\left(\frac{2R}{\gamma}\right)^2$



Ejercicio

- Cuales son los valores de w , b y k para la función AND? En la forma primal
- Cuales son los valores de α , b y k para la función AND en forma dual

