

Aprendizaje de Maquina y Reconocimiento de Patrones

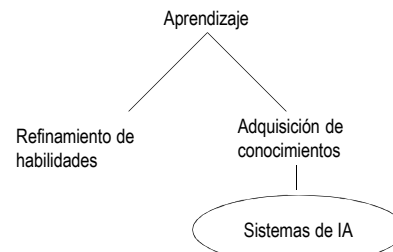
Inteligencia x Aprendizaje

- El aprendizaje es la llave de la superioridad de la Inteligencia humana
 - El aprendizaje es la esencia de la inteligencia
- Para que la maquina tenga comportamiento inteligente se debe aumentar la capacidad de aprendizaje

Aprendizaje y Adaptación

- Modificación de una tendencia comportamental por medio de experiencia
- Una maquina de aprendizaje definida de manera amplia es cualquier dispositivo cuyas acciones son influenciadas por experiencias anteriores
- Cualquier cambio en un sistema permite tener un mejor desempeño en vez de que el repita una misma tarea u otra reiterada en la misma población

Aprendizaje



Aprendizaje

- Definición
 - Es una sub área de la IA que investiga métodos computacionales relacionados a la adquisición de nuevos conocimientos relacionados a nuevas habilidades y nuevas formas de organizar el conocimiento existente
 - Se dice de un programa computacional aprende a partir de la experiencia E en relación a una clase de tareas T con la medida de desempeño P si su desempeño en las tareas T , medida por P mejora con la experiencia E

Inteligencia x Aprendizaje

- El ser humano esta pre-programado para el aprendizaje. Aprende ampliando el alcance del conocimiento que ya posee a través de reordenaciones
- El computador no posee el programa inicial para buscar por informaciones y realizar aprendizaje en general

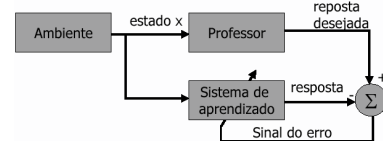
Objetivos del AM

- Un mejor entendimiento de los mecanismos de aprendizaje humano

Automatización del proceso de adquisición del conocimiento.

Aprender con un Profesor

- Conocido como aprendizaje supervisado
- El conocimiento es representado por un conjunto de valores de entrada y salida
- Minimiza el error en función a las respuesta obtenida con la deseada

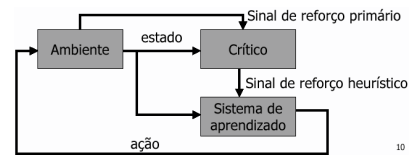


Cursos Relevantes

- Inteligencia Artificial
- Métodos Bayesianos
- Teoría de control
- Teoría de la información
- Teoría de la complejidad computacional
- Filosofía
- Psicología y Neurobiología
- Estadística

Aprender con un Critico

- Aprender por medio de interacciones con el ambiente
- Explora estado de acciones
- *Feedback* por medio de señales de refuerzo
- Objetivo maximizar la acumulación futura de refuerzo

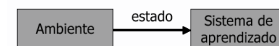


Aprendizaje

- Paradigmas del Aprendizaje
 - Aprendizaje con un profesor
 - Aprendizaje con un crítico
 - Aprendizaje no supervisado
- Tareas del Aprendizaje
 - Asociación de patrones
 - Agrupamiento
 - Reconocimiento de patrones
 - Aproximación de patrones
 - Control
 - Filtros

Aprendizaje no Supervisado

- Aprendizaje auto organizable
- No hay un profesor o crítico
- Medida de calidad dependerá de la tarea
- Descubrirá agrupamientos automáticamente
- Aprendizaje competitivo



Asociación de Patrones

- Técnica que asocia un dato patrón a una salida
 - Hetero asociativa
 - Auto asociativa
 - Asociarse consigo mismo
- Ejemplos
 - Asociar imágenes de rostros con personas
 - Recuperación de imágenes por el contenido

Clasificación

- Técnica que clasifica nuevas entradas en una de varias categorías discretas
 - Numero definido de categorías
 - Clasificación jerárquica
- Ejemplos
 - Asignación de Crédito
 - Reconocimiento de objetos

Agrupamiento

- Técnica que explora las semejanzas entre patrones y agrupar patrones parecidos.
- También se le conoce como aprendizaje no supervisado
- Análisis de datos
 - Extrae información de un conjunto de datos
- Ejemplos
 - Minería de datos
 - Compresión de datos

Regresión

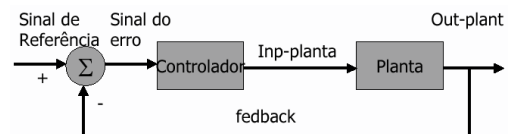
- Técnica que asocia nuevas entradas a un valor continuo
 - Aprende una función definida en términos de una medida sobre una cantidad aleatoria
 - Las salidas son variables continuas
- Ejemplos
 - Previsión de series temporales
 - Temperatura
 - Caudal de los ríos

Aproximación de Funciones

- Encontrar una estimativa F' de una función desconocida F
 - Se conoce el conjunto de pares de entrada
 - $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$
- Puede ser:
 - Clasificación
 - Regresión

Control

- Ajustar parámetros de un controlador tal que el sistema controle el *loop* cerrado de un comportamiento



Formas de realizar AM

- Simbólico
- *Instance Based*
- Estadístico
- Conexionista
- Genético

AM Estadístico

- Como regla general, técnicas estadísticas tienden a focalizar tareas en que todos los atributos tienen valores continuos u ordinales. Muchos de ellos también son parámetros, asumiendo alguna forma de modelo y entonces encontrando valores apropiados para parámetros del modelo a partir de datos.
- Por ejemplo un clasificador lineal asume que las clases pueden ser expresadas como una combinación lineal particular que provee una mejor aproximación sobre un conjunto de datos.

AM Simbólico

- Técnica que explora representaciones de estructuras gráficas o lógicas en lugar de métodos estadísticos numéricos
- Sistemas de AM simbólico aprende descripciones simbólicas de alto nivel, haciendo fuertes suposiciones estructurales sobre los conceptos a ser adquiridos

AM Conexionista

- Redes neuronales son construcciones matemáticas relativamente simple que fueron inspiradas en el sistema nervioso
- Su representación envuelve un conjunto de unidades interconectadas en el cual el nombre conexionismo es usado para describir el área de estudio

AM Instance Based

- Una forma de clasificar un caso y recordar uno similar cuya clase es conocida y se asume que el nuevo caso tendrá la misma clase
- Esta filosofía ejemplifica los sistemas *Instance-based* que clasifican casos nunca vistos a través de casos similares

AM basado en la Genética

- Un clasificador genético consiste de una población de elementos de clasificación que compiten para hacer una predicción. Elementos que poseen una performance débil serán descartados, mientras que los elementos mas fuertes proliferan, produciendo variaciones entre ellos
- Este paradigma posee analogía directa con la teoría de Darwin, donde sobreviven los mas aptos para el ambiente donde se encuentran

Validación de Sistemas de AM

- Experimental,
 - Conducir experimentos con validación cruzada
 - Comparar varias técnicas usando varios conjuntos de datos
 - Obtener datos de sus experimentos
 - Precisión de prueba
 - Tiempo de entrenamiento
 - Tiempo de prueba
 - Analizar las diferencias
 - Media
 - Desviación estándar

Próximos Pasos

- Mañana: impacto enorme
 - Aprender utilizando mezclas de datos de diferentes tipos
 - Utilizar múltiples bases de datos internas, además de la web
 - Aprender por experimentación activa
 - Aprender decisiones en vez de previsiones
 - Aprendizaje acumulativo, de larga duración
 - Lenguajes de programación con aprendizaje incluido

Validación de Sistemas de AM

- Teórica,
 - Analizar algoritmos matemáticamente y probar los teoremas sobre:
 - Complejidad computacional
 - Habilidad para ajustar datos de entrenamiento
 - Complejidad de la muestra
 - Numero de ejemplos de entrenamiento necesarios para aprender una función con una precisión dada

Preguntas en AM

- Que algoritmos pueden aproximar funciones y cuando usarlos?
- Como el número de ejemplos de entrenamiento influyen la precisión?
- Como el ruido de los datos influencia la precisión obtenida?
- Como el conocimiento a priori puede auxiliar el aprendizaje?
- Que inspiraciones pueden ser obtenidas de sistemas de aprendizaje biológico?
- Como los sistemas pueden alterar su propia representación?
- Cuales son los límites teóricos de lo que es posible aprender?

Estado Actual

- Resultados teóricos han sido desarrollados caracterizando la relación entre:
 - Número de ejemplos de entrenamiento observados
 - Número de hipótesis bajo consideración
 - Error esperado en las hipótesis aprendidas
 - Comenzamos a:
 - Obtener modelos iniciales del aprendizaje humano y animal
 - Entender su relación con los algoritmos de aprendizaje desarrollados para computadoras
- Hoy: es la punta del iceberg
 - Algoritmos de la primera generación: redes neuronales, árboles de decisión, regresión aplicados a bases de datos bien definidas

Validación de Sistemas de AM

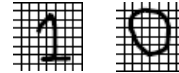
- Machine Learning at Google
 - http://directory.google.com/Top/Computers/Artificial_Intelligence/
- Matlab Toolbox for Pattern Recognition
 - <http://www.ph.tn.tudelft.nl/~bob/PRTOOLS.html>
- MIT GALIB in C++
 - <http://lancet.mit.edu/ga>
- Machine Learning Data Repository UC Irvine
 - <http://www.ics.uci.edu/~mlearn/ML/Repository.html>

Reconocimiento de Patrones

Ejemplo



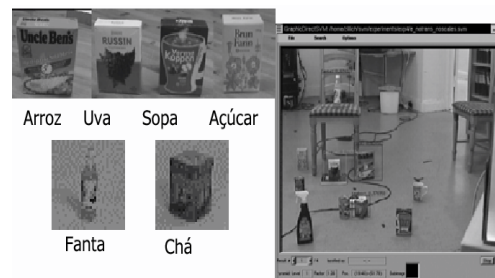
- Reconocimiento de caracteres
 - Cada carácter esta formado por varias variables
 - $X = \{x_1, x_2, x_3, \dots, x_n\}$



Reconocimiento de Patrones

Introducción

- Un de los aspectos mas interesantes del mundo es que el puede estar hecho de patrones. Un patrón es esencialmente un arreglo que esta caracterizado por el orden de los elementos de los cuales el esta hecho al contrario de la naturaleza intrínseca de estos elementos
Norbert Wiener
- Las redes neuronales artificiales y el reconocimiento de patrones atacan los mismos problemas



Reconocimiento de Patrones

- Definición: técnica de toma de decisiones a partir de patrones complejos de informaciones
- Objetivo: desarrollar sistemas capaces de lidiar con tareas realizadas por seres humanos, como el reconocimiento.
- Función: clasificación de patrones de entrada en categorías o clases

Reconocimiento de patrones

- Objetivo: reconocer objetos en una imagen
- Entrada: imagen
- Decisión: saber si contiene el objeto o no
- Posibles características relevantes que ayudan a definir una respuesta
 - Datos de la imagen
 - Histogramas de color
 - Filtros espaciales
 - Contornos



Problema de la Dimensionalidad

- Usar todos los valores del vector de características es un problema ya que en muchos casos la dimensión es muy alta
 - Imágenes de 256 * 256 pixels de resolución generan vectores de tamaño 65 536
 - Datos del genoma humano también son de alta dimensión
- Muchas de las características usadas están de mas
- La cantidad de patrones es otra limitante.

Extracción de características

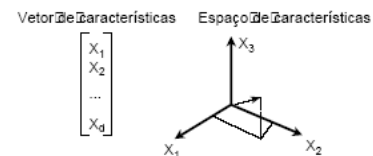
- Puesto que uno de los problemas es la dimensionalidad la extracción de características ayuda reduciendo la dimensión a través de diferentes técnicas
 - PCA (análisis de componentes principales)
 - Transformadas (Fourier wavelets, etc)
 - Proyecciones

Grados de Libertad



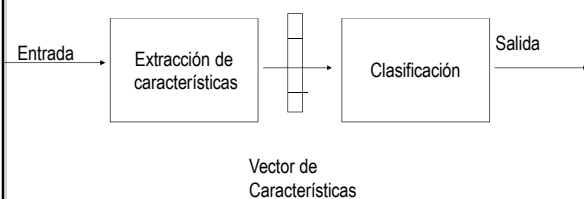
Vector de Características

- El vector de características es el objeto o evento que puede ser representado abstractamente como un punto en un espacio de características



Etapas del R. P.

- Extracción de Características
- Clasificación



Selección de Características

- Elegir aquellas características que sean relevantes para la clasificación
 - Uno de los problemas es no saber cuales de esas características son las mas relevantes
 - Muchas de esas características son identificas en el proceso de clasificación
- Un buen conjunto de características es generalmente un proceso empírico pero existen algunas heurísticas que nos ayudan

Selección de Características

- Principales Heurísticas
 - Método exhaustivo
 - Selección paso a paso
 - Combinación de características

Selección paso a paso

- Evita consideraciones de todos los subconjuntos
 - Selección forward
 - Comienza seleccionando la mejor característica
 - Después adiciona mas características que mejoran el desempeño
 - Eliminación backward
 - Borra la característica que reduce el desempeño

Método Exhaustivo

- Procedimiento también llamado de fuerza bruta
 - Dividir las d características en subconjuntos de m características
 - Usar un clasificador para cada subconjunto
 - Usar datos independientes para estimar la tasa de errores de cada clasificador
 - Usar el subconjunto con menor tasa de error

Combinación de características

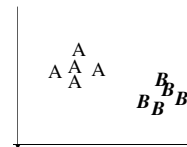
- Combina las características originales para generar un conjunto menor, mas efectivo
 - Sea X el vector de características original de dimensión d y W de dimensión $d \times m$ ($m < d$)
 - El nuevo vector de características Y , de dimensión m es definido por $Y = W^T X$
 - El problema es definir la matriz W
- Posibles soluciones al problema
 - Análisis de componentes principales PCA
 - Análisis de componentes múltiples MCA

Problemas

- Útil apenas para un conjunto de datos pequeño
- Existe una manera muy diferente de encontrar separar los subconjuntos
- Las características puede ser buenas solo para la prueba y no para un entrenamiento
- Los resultados dependen de m

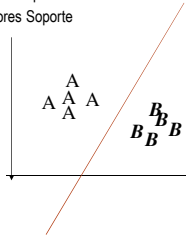
Clasificación

- El objetivo es aprender una función que mapea un ejemplo en una clase
- Ejemplos



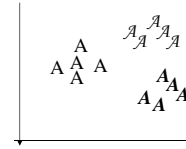
Clasificación

- Separación lineal
 - Para separa se puede usar
 - Red Neuronal Perceptron
 - Maquinas Vectores Soporte



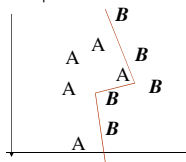
Clustering

- Frecuentemente patrones de una misma clase no son homogéneos
 - Esta compuesta de subclases distintas
 - Un clasificador puede considerar cada subclase como una categoría diferente
 - Se puede usar el algoritmo k-medias



Clasificación

- Separación no lineal
 - Para separa puede usarse:
 - Red neuronal multicapas perceptron (BackPropagation)
 - Maquinas Vectores Soporte



Clustering K - medias

- Paso 1: Escoger k centros iniciales de grupos que pueden ser arbitrariamente escogidos de un conjunto de vectores $s = \{x_1, x_2, \dots, x_n\}$ a ser clasificados de un conjunto de n de esa manera se asigna un conjunto de centros k la cual es un parámetro el conjunto de centros es:

$$k = \{z_1(1), z_2(1), \dots, z_n(1)\}$$

Clasificadores

- Existen dos maneras de proyectar un clasificador
 - Crear un modelo matemático del problema y derivar un clasificador optimo
 - Ajusta una posible solución y ajustarla para encajar al problema
 - Recibir un entrada
 - Definir clases para la entrada
 - Posibles clasificaciones
 - Correcta
 - Incorrecta
 - Tolerancia de Error
- Las redes neuronales es unas de las técnicas mas usadas en clasificación.

K - medias

- Paso 2 Distribuir los objetos de $x \in s$ entre los k grupos $s_j(t)$ asumiendo el criterio de:

$$x \in S_j(t) \text{ se } \|x - z_j(t)\| < \|x - z_i(t)\|,$$

$$\forall i, j = 1, 2, \dots, K \text{ e } j \neq i$$

K - medias

- Paso 3: Como los centros fueron arbitrariamente distribuidos, se calculan los nuevo centros como la media de cada grupo así el de $z(t)$ asumirá nuevos valores dados por:

- Donde N es el numero de vectores que pertenecen a $s_j(t)$
- $$z_j(t+1) = \frac{1}{N_j} \sum_{x \in S_j(t)} x$$

K- medias ejemplo

- Paso 1: $k = 2$, escoger $z_1() = x_1 = (0,0)$
 $z_2() = x_2 = (0,1)$

- Paso 2:

$i = 2$, entonces $s_1 = \{x_1, x_3\}$ y el resto pertenece a $s_2 =$
 $\|x_1 - z_1(1)\| < \|x_1 - z_1(1)\|$ e $\|x_3 - z_1(1)\| < \|x_3 - z_1(1)\|$.

K - medias

- Paso 4: probar la convergencia de los nuevos centros:

$$z(t) = z(t+1)$$

si son diferentes volver al Paso 2

si son iguales entonces termino

K – medias ejemplo

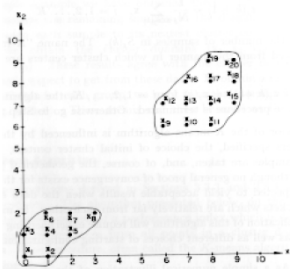
- Paso 3 : Actualizar los centros del cluster

$$z_1(2) = \frac{1}{N_1} \sum_{x \in S_1(t)} x = \frac{1}{2} (x_1 + x_3) = \begin{pmatrix} 0.0 \\ 0.5 \end{pmatrix}$$

- Pa: $z_2(2) = \frac{1}{N_2} \sum_{x \in S_2(t)} x = \frac{1}{18} (x_2 + x_4 + \dots + x_{20}) = \begin{pmatrix} 5.67 \\ 5.33 \end{pmatrix}$

$$z_j(2) \neq z_j(1), j = 1, 2,$$

K – medias ejemplo



Métricas

- Distancia de Hamming
- Distancia Euclideana
- Distancia Manhattan
- Distancia Cuadratica

Medidas Distancia

- Distancia de Hamming
 - Distancia usada para vectores binarios

$$D_H(X, Y) = \sum_{i=1}^N |x_i - y_i|$$

- Distancia Euclídeana

$$D_E(X, Y) = \sqrt{\sum_{i=1}^N (x_i - y_i)^2}$$

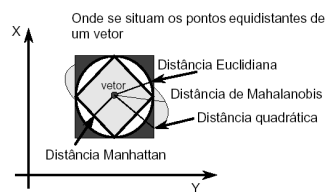
- Distancia de Manhattan

$$D_{bc}(X, Y) = \sum_{i=1}^N |x_i - y_i|$$

Medidas de Distancia

- Distancia Cuadrática

$$D_q(X, Y) = \text{MAX}(|x_i - y_i|)$$



Reconocimiento de Patrones

- En la actualidad es un area de estudio muy aplicada por ejemplo
 - En Biometría
 - Reconocimiento de rostros
 - Reconocimiento de huellas digitales
 - En Bioinformática
 - Alineamiento de cadenas de ADN
 - En Hidrología
 - Predicción del caudal de los ríos
 - En Medicina
 - Determinación de tipos de parásitos
 - Helmintos