Tópicos Selectos en Aprendizaje Maquinal

Guía de Trabajos Prácticos N°1

Algoritmos para Reconocimiento de Patrones

29 de octubre de 2018

1. Objetivos

- Introducir conceptos básicos de aprendizaje automático.
- Conocer los principales aspectos de implementación práctica de diversos algoritmos de reconocimiento de patrones.
- Estudiar y validar su comportamiento con algunos ejercicios de "lápiz y papel" y sobre conjuntos de datos artificiales sencillos.

2. Ejercicios

Ejercicio 1: Introducción y repaso de probabilidad y teoría de información.

Escriba un programa para generar números aleatorios $\mathbf{x}$ con una función de densidad de probabilidad (*fdp*) gaussiana $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ en d dimensiones[1]:

1. Para cada uno de los siguientes casos: i) *fdp* gaussiana unidimensional; ii) *fdp* gaussiana bidimensional no correlacionada y iii) *fdp* gaussiana bidimensional correlacionada,
   a) Genere un conjunto de números aleatorios con esa *fdp*.
   b) Estime la *fdp* experimental mediante un histrograma normalizado.
   c) Compare graficamente la estimación obtenida con la *fdp* teórica.
2. Compruebe numéricamente el teorema del límite central mediante la suma de números aleatorios con distribución uniforme.
3. Estime numéricamente y compare con el valor teórico, la entropía de una variable aleatoria con distribución: i) laplaciana; ii) gaussiana y iii) uniforme.

[1] Pueden realizarse varios programas, avanzando gradualmente en complejidad.

Ejercicio 2: Clasificación estadística de patrones.

1. Sea un clasificador geométrico lineal definido por:
   $g_1(\mathbf{x}) = -x_1$
   $g_2(\mathbf{x}) = x_1 + x_2 - 1$
   $g_3(\mathbf{x}) = x_1 - x_2 - 1$

   a) Calcule y grafique las fronteras y regiones de decisión.
   b) Clasifique los puntos: (2,1); (2,-1); (-1,1); (-1,-1).
2. Sean A y B dos clases de igual probabilidad a priori definidas por:
   $P(\mathbf{x}|A) \rightsquigarrow \mathcal{N}\left(\binom{3}{0}, \left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)\right)$
   $P(\mathbf{x}|B) \rightsquigarrow \mathcal{N}\left(\binom{0}{3}, \left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)\right)$

   a) Construya un clasificador gaussiano y encuentre la frontera.
   b) Realice una representación gráfica del problema.
3. Implemente un clasificador gaussiano por ML para el mini-corpus de muestras proporcionado, realizando una representación gráfica de la situación:
   a) Clasifique los datos del conjunto de entrenamiento y calcule la tasa de aciertos.
   b) A partir de los parámetros obtenidos genere *nuevos* datos de prueba, clasifíquelos y compare con el resultado anterior.

Ejercicio 3: Análisis estadístico de datos - PCA.

1. Implemente el algoritmo de PCA.
2. Escriba un programa que le permita generar datos aleatorios $\mathbf{x}$ a partir del siguiente modelo generativo lineal: $\mathbf{x} = \mathbf{A}\mathbf{s}$, donde $\mathbf{s}$ es el vector de fuentes (aleatorio) y $\mathbf{A}$ es la matriz de mezcla.
3. A partir de datos de dos mezclas, obtenidos mediante dos fuentes y una matriz de mezcla aleatoria, utilice PCA para lo siguiente:
   a) Pruebe con fuentes con distribución gaussiana y laplaciana, para matrices de mezcla con columnas ortogonales y no ortogonales (i.e. $\left(\begin{smallmatrix}+0{,}2 & +1\\ +0{,}2 & -1\end{smallmatrix}\right)$ y $\left(\begin{smallmatrix}+0{,}6 & +0{,}8\\ +0{,}3 & -0{,}2\end{smallmatrix}\right)$, respectivamente):
   b) Para cada caso de los anteriores y cada etapa (fuentes, mezclas, señales separadas) dibuje un gráfico de dispersión de las variables.
   c) Luego de la separación, obtenga la matriz $\mathbf{W}$ correspondiente.
   d) ¿La matriz de separación $\mathbf{W}$ es la inversa de la matriz de mezcla $\mathbf{A}$ utilizada?
   e) ¿Cómo se afecta este resultado si agrega una componente de ruido gaussiano al modelo generativo?

Ejercicio 4: Análisis estadístico de datos - ICA.

1. Implemente el algoritmo de FastICA con aprendizaje deflacionario.
2. A partir de datos de dos mezclas, obtenidos mediante dos fuentes laplacianas y una matriz de mezcla aleatoria, utilice FastICA para lo siguiente:
   a) Para cada etapa (fuentes, mezclas, señales blanqueadas, señales separadas) dibuje un gráfico de dispersión de las variables.
   b) Luego de la separación, estime las matrices $\mathbf{P}$ y $\mathbf{D}$.
   c) ¿La matriz de separación $\mathbf{W}$ es la inversa de la matriz de mezcla $\mathbf{A}$ utilizada?
   d) ¿Cómo se afecta este resultado si agrega una componente de ruido gaussiano al modelo generativo?
   e) ¿Que ocurre si una de las fuentes es gaussiana?¿Y si ambas fuentes lo son?

Ejercicio 5: Aprendizaje basado en redes y algoritmos clásicos

1. Genere un conjunto de datos aleatorios bidimensionales separados en dos clases gaussianas correlacionadas con cierto grado de solapamiento.
2. Implemente el algoritmo de entrenamiento de un perceptron simple y pruébelo con los datos anteriores, grafique los resultados obtenidos.
3. Implemente el algoritmo de aprendizaje no supervisado k-medias (batch) y pruébelo con los datos anteriores, grafique los resultados obtenidos.
4. Implemente el algoritmo de aprendizaje de las redes neuronales con funciones de base radial y pruébelo con los datos anteriores, grafique los resultados obtenidos.
5. Implemente el algoritmo de los k-vecinos más cercanos y pruébelo con los datos anteriores, grafique los resultados obtenidos.
6. Compare y comente los resultados obtenidos por los distintos métodos.

Ejercicio 6: Redes neuronales dinámicas.

1. Implemente la arquitectura y entrenamiento Hebbiano para una red recurrente de Hopfield con los patrones que se muestran en la Figura 1.
2. Utilice la red de Hopfield como memoria asociativa para los patrones de la Figura 2.
3. Agregue diferentes cantidades de ruido a los patrones de la Figura 2 y utilice estos ejemplos ruidosos para acceder a las memorias fundamentales almacenadas como en el ejercicio anterior. Para simular cantidades controladas de ruido se sugiere invertir (blanco a negro y viceversa) cada pixel del dígito con probabilidades 0.1, 0.2 y 0.5.

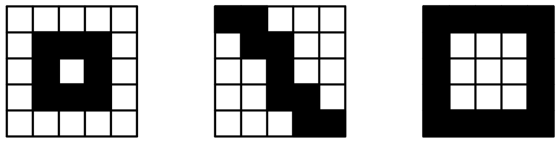


Figura 1: Patrones sencillos para almacenar en las memorias fundamentales.

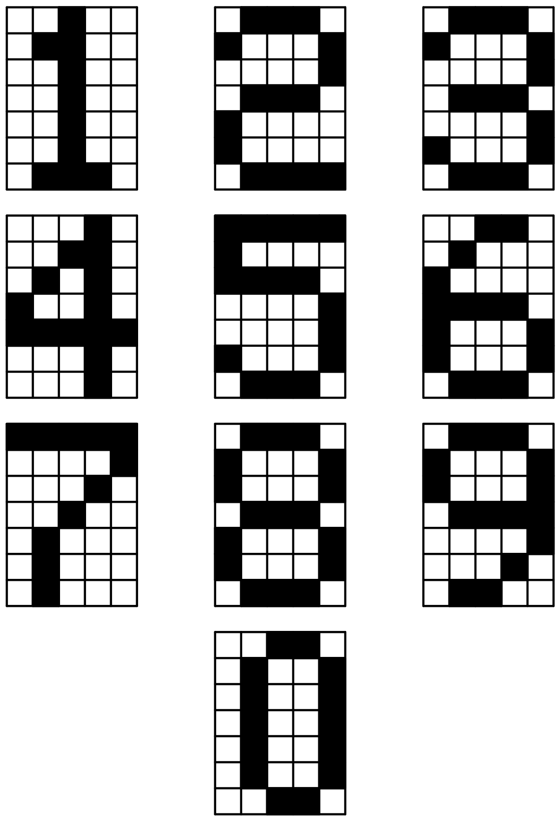


Figura 2: Memorias fundamentales hechas a partir de los dígitos de 0 a 9.

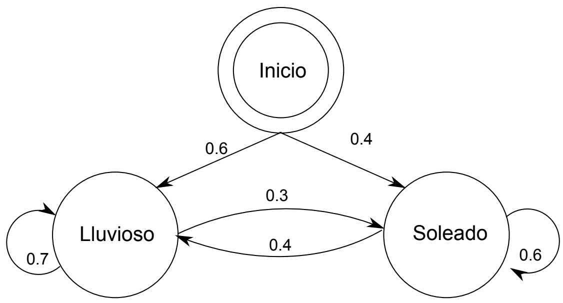


Figura 3: Cadena de Markov de 2 estados para modelar el clima diario en una ciudad.

Ejercicio 7: Modelos de Markov.

1. La Secretaria de Turismo de una ciudad quiere modelar el clima diario local para planificar distintas actividades. Para ello se propone el modelo probabilístico de la Figura 3 basado en autómatas observables de Markov, con los siguientes parámetros:

$$A = \begin{bmatrix} 0{,}7 & 0{,}3 \\ 0{,}4 & 0{,}6 \end{bmatrix}$$

$$\pi = \begin{bmatrix} 0{,}6 & 0{,}4 \end{bmatrix}$$

   a) Realice la simulación computacional del mismo como modelo generativo.
   b) Genere varias secuencias climáticas y grafíquelas.
   c) Utilice las secuencias de salida generadas para re-estimar los valores de A y π: ¿Cómo influye la cantidad de secuencias y el largo de las mismas en la estimación de los parámetros del autómata? ¿Sería este modelo válido para todas las estaciones del año?
   d) ¿Cómo convertiría este modelo en un modelo oculto de Markov? Realice los cambios que correspondan en la simulación para este caso, genere nuevas secuencias con este modelo y comente los resultados comparados con el caso anterior[2].

2. La Secretaria de Turismo quiere ahora modelar la actividad principal diaria de los turistas que visitan la ciudad y su relación con el clima. Para ello se propone completar el modelo anterior para convertirlo en un modelo oculto de Markov como el de la Figura 4, con los siguientes parámetros:

$$A = \begin{bmatrix} 0{,}7 & 0{,}3 \\ 0{,}4 & 0{,}6 \end{bmatrix}$$

[2] Puede comenzar suponiendo una matriz de probabilidades de observación tipo identidad (caso completamente observable) para luego "ocultar" progresivamente el modelo.

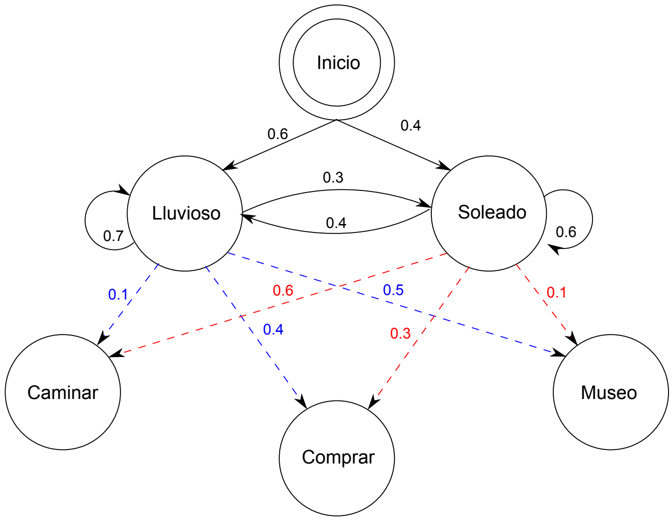


Figura 4: Modelo oculto de Markov de 2 estados y 3 salidas posibles para modelar la actividad de un turista.

$$B = \begin{bmatrix} 0{,}1 & 0{,}6 \\ 0{,}4 & 0{,}3 \\ 0{,}5 & 0{,}1 \end{bmatrix}$$

$$\pi = \begin{bmatrix} 0{,}6 & 0{,}4 \end{bmatrix}$$

a) Genere una secuencia de comportamientos suficientemente larga y calcule la secuencia de climas más probable (para esa secuencia de comportamientos). ¿Cuánto difiere la secuencia climática "real" generada por el modelo de la estimada a partir de sus parámetros?

b) Utilice el modelo para analizar el desempeño de los métodos de entrenamiento:

- Genere varias secuencias adicionales de comportamiento.
- A partir de estos datos estime los parámetros del modelo mediante el algoritmo de Viterbi y el de Baum-Welch.
- Compare los resultados obtenidos en ambos casos con los valores reales de los parámetros.

c) Si tuviera varios modelos que describieran distintos hábitos de comportamiento para distintos tipos de turistas: ¿cómo implementaría un método que le permita clasificar un turista recién llegado en función de su secuencia de actividades diaria durante la primer semana de estadía?

Ejercicio 8: Árboles de decisión.

1. Se tienen datos acerca de la realización o suspensión de partidos de tenis en función del pronóstico del tiempo y los datos del día que se han volcado en el Cuadro 1:

Cuadro 1: Datos de juegos de tenis en función del pronóstico del tiempo.

#	Atributo				Clase
	Pronóstico	Temp.	Humedad	Viento	Juega
1	soleado	calor	alta	no	N
2	soleado	calor	alta	si	N
3	nublado	calor	alta	no	P
4	lluvioso	moderado	alta	no	P
5	lluvioso	frío	normal	no	P
6	lluvioso	frío	normal	si	N
7	nublado	frío	normal	si	P
8	soleado	moderado	alta	no	N
9	soleado	frío	normal	no	P
10	lluvioso	moderado	normal	no	P
11	soleado	moderado	normal	si	P
12	nublado	moderado	alta	si	P
13	nublado	calor	normal	no	P
14	lluvioso	moderado	alta	si	N

Cuadro 2: Datos de juegos de tenis en función del pronóstico del tiempo con atributos numéricos.

#	Atributo				Clase
	Pronóstico	Temp.	Humedad	Viento	Juega
1	soleado	85	85	no	N
2	soleado	80	90	si	N
3	nublado	83	78	no	P
4	lluvioso	70	96	no	P
5	lluvioso	68	80	no	P
6	lluvioso	65	70	si	N
7	nublado	64	65	si	P
8	soleado	72	95	no	N
9	soleado	69	70	no	P
10	lluvioso	75	80	no	P
11	soleado	75	70	si	P
12	nublado	72	90	si	P
13	nublado	81	75	no	P
14	lluvioso	71	80	si	N

a) Construya “a mano”(haciendo todas las cuentas) y dibuje el árbol binario de decisión que describa los datos sobre juegos de tenis del cuadro. Utilice como impureza la entropía.

b) Escriba un programa que permita crecer árboles de decisión binarios y utilícelo para generar automáticamente un árbol a partir de los datos anteriores. Contraste con el árbol generado por Ud.

c) Modifique el tipo de impureza empleada y comente las diferencias con el árbol generado en el punto anterior.

2. Se han modificado los datos de juegos de tenis anteriores para incluir valores numéricos de temperatura y humedad de acuerdo con el Cuadro 2: Modifique el programa desarrollado para tratar con atributos numéricos y genere un nuevo árbol con los datos del cuadro.

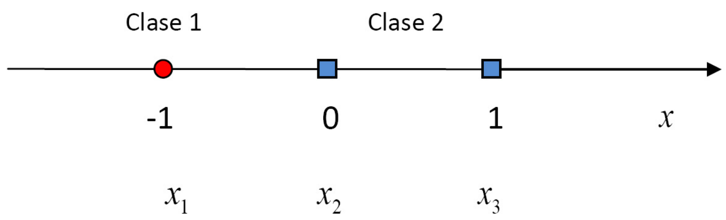


Figura 5: Problema unidimensional linealmente separable.

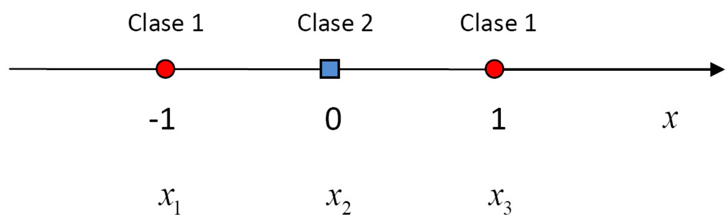


Figura 6: Problema unidimensional no linealmente separable.

Ejercicio 9: Máquinas de vectores soporte

1. Considere el problema unidimensional linealmente separable de la Fig. 5: Se pide:
   a) Escriba todas las inecuaciones de las restricciones.
   b) Plantee el problema dual de optimización.
   c) Encuentre los valores de α_i para $i = 1, 2, 3$.
   d) Encuentre la función de decisión y los vectores soporte. Grafique.
2. Considere el problema unidimensional no linealmente separable de la Fig. 6: Utilice un núcleo dado por un polinomio de segundo grado del tipo[3]:

$$k(\mathbf{x}, \mathbf{x}') = (\mathbf{x} \cdot \mathbf{x}' + 1)^2.$$

   Se pide:
   a) Plantee el problema dual de optimización.
   b) Encuentre los valores de C y de α_i para $i = 1, 2, 3$.
   c) Encuentre la función de decisión y los vectores soporte. Grafique.

Ejercicio 10: Aprendizaje por refuerzo

Para estudiar y validar el comportamiento de los algoritmos de aprendizaje por refuerzo pasivo se proponen los siguientes ejercicios de lápiz y papel en un "problema de juguete":

1. Encontrar la política óptima para un agente pasivo en el mundo de 4x3 de la Figura 7, con recompensa -0.04 en los estados no terminales. En la Fig. 8 se pueden observar ejemplos de políticas posibles para este problema.

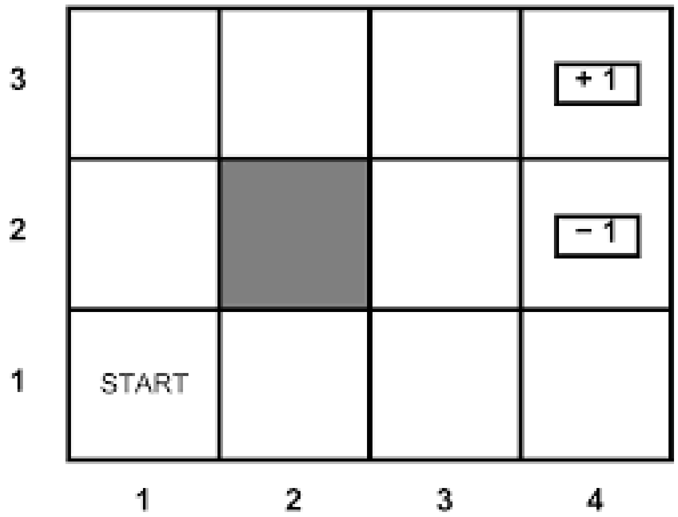


Figura 7: Mundo tipo grilla de 4x3.

$(1,1)_{-.04} \rightarrow (1,2)_{-.04} \rightarrow (1,3)_{-.04} \rightarrow (1,2)_{-.04} \rightarrow (1,3)_{-.04} \rightarrow (2,3)_{-.04} \rightarrow (3,3)_{-.04} \rightarrow (4,3)_{+1}$

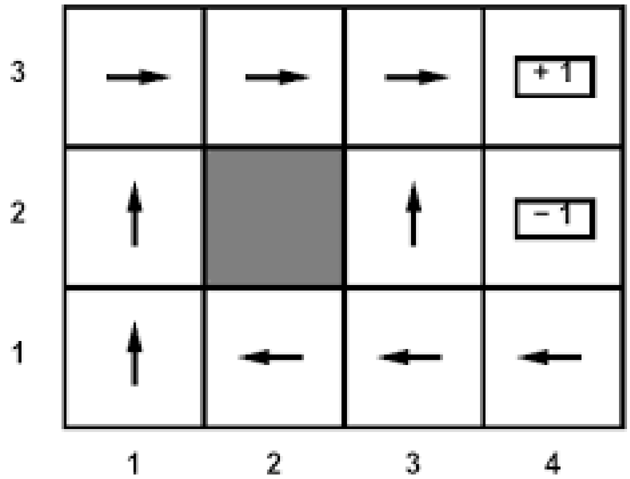
$(1,1)_{-.04} \rightarrow (2,1)_{-.04} \rightarrow (3,1)_{-.04} \rightarrow (3,2)_{-.04} \rightarrow (4,2)_{-1}$

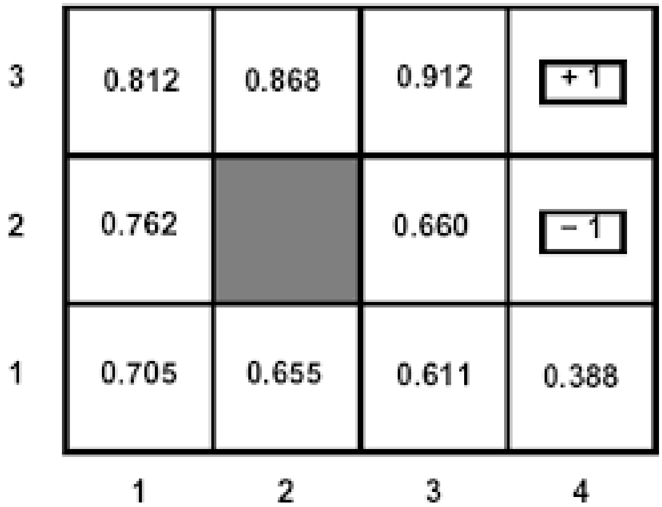
Figura 8: Algunas políticas posibles para el mundo de 4x3.

2. Calcular las utilidades de los estados en el mundo de 4x3, dada la política óptima encontrada en el ejercicio anterior.

[3] Se agrega el 1 de forma que todos los términos cruzados con grados iguales o menores que 2 están incluidos.

Respuestas:





Referencias

[1] C.M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, 2006.

[2] S. Kay. *Intuitive Probability and Random Processes using MATLAB®*. Intuitive Probability and Random Processes Using MATLAB. Springer US, 2006.

[3] Stuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson Education, 2 edition, 2003.