

Reconocimiento de patrones

Clase 8: clasificador de vecino más cercano



iimas



Para el día de hoy...

- Anteproyecto
- Clasificador del vecino más cercano



Entrega de anteproyecto (2 semanas)

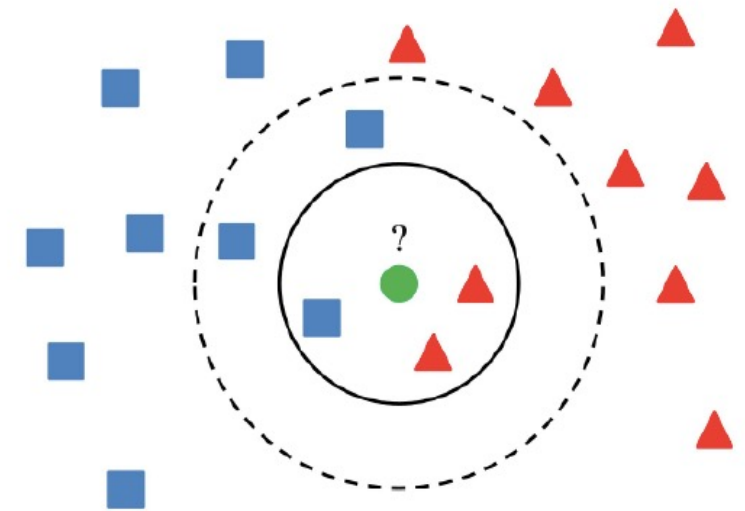
- Entregar en un documento pdf de una página los siguientes elementos para definir el proyecto final del curso:
 - Título
 - Miembros del equipo (max 3)
 - Descripción del problema¿Qué no está sucediendo?
- ¿Quién está interesado?
- ¿Por qué es de interés para ustedes?
- ¿Qué previene que se haga?
- ¿Qué pasa si no se hace?
- ¿Qué se planea hacer? (en términos de lo existente y las operaciones básicas para transformar)
- ¿Cómo luce el éxito?

Clasificador multiprototipo

- Dado un conjunto de clases C_i con un conjunto de prototipos y_j
- Dado un nuevo ejemplo x
 - Calculamos el prototipo más cercano de cada clase
 - Asignamos x a la clase que tenga el prototipo más cercano a x
- ¿Cómo elegimos los prototipos de cada clase?

Clasificación de vecino más cercano (NN)

- Consideremos el caso de m clases $\{C_i\}_{i=1}^m$ y un conjunto de patrones $\{y_i\}_{i=1}^N$ clasificados a priori
- Sea x una nueva observación
- El clasificador de vecino más cercano clasifica x en la clase de su vecino más cercano en el conjunto $\{y_i\}_{i=1}^N$
- Si $\|x - y_j\| = \min_{1 \leq i \leq N} \|x - y_i\|$ entonces $x \in C_j$
- El enfoque se puede modificar para considerar los k vecinos más cercanos



Ejercicio

- Considere un problema de dos clases en \mathbb{R}^2 . Suponga que los siguientes patrones han sido clasificados a priori
 - $(1,1), (2,3), (2,1), (2,2) \in C_1$
 - $(4,0), (3,-1), (3,1) \in C_2$
- Clasifique $x = (2.2, 0)$
 - De acuerdo a 1-NN
 - De acuerdo a 3-NN

Algoritmo

- Entradas: n las dimensiones del problema, N el número de patrones preclasificados, m el número de clases, (x_i, j_i) , $1 \leq i \leq N$ pares ordenados del patrón y su clase, k orden del clasificador NN, x un nuevo patrón a clasificar
- Salida: l la clase asignada a x
- $S = \{x_i, j_i\}_{i=1}^N$
- Encontrar $(y, j_0) \in S$ que satisface $\|y - x\| = \min \|z - x\|, (z, j) \in S$
- Si $k = 1$, $l = j_0$ y finalizar
- Si no $IC(i') = 0, i' \neq j_0, IC(j_0) = 1$ y $S = S - \{(y, j_0)\}$
- Desde $i_0 = 1, \dots, k - 1$
 - Encontrar $(y, j_0) \in S$ que satisface $\|y - x\| = \min \|z - x\|, (z, j) \in S$
 - $IC(j_0) = IC(j_0) + 1$ y $S = S - \{(y, j_0)\}$
 - $l = \max\{IC(i)\}, 1 \leq i \leq m$

Análisis del algoritmo

¿Es correcto?

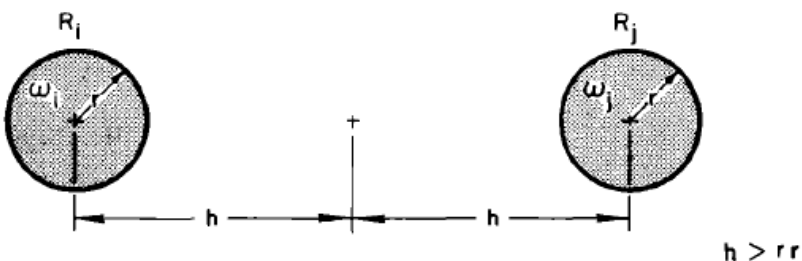
¿Complejidad de tiempo?

¿Complejidad de espacio?



¿Cuál k es
mejor?





El error para 1-NN

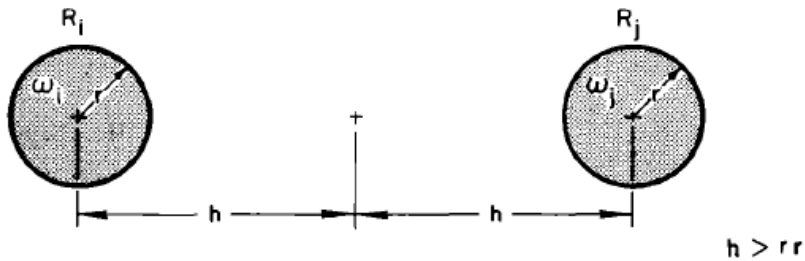
- Supongamos que dos clases C_i y C_j tienen la misma probabilidad de ocurrir
- Las clases están distribuidas uniformemente sobre los discos R_i y R_j
- Para N muestras la probabilidad de obtener exactamente α muestras de la clase C_i está dada por
- $p_i = \frac{1}{2^N} C_\alpha^N$
- Donde $C_\alpha^N = \frac{N!}{\alpha!(N-\alpha)!}$
- Suponga $x \in C_i$ pero 1-NN lo clasifica en C_j
- Para 1-NN $\alpha = N$ y $p_{e_1} = \frac{1}{2^N}$

El error para k-NN

- Suponga que $x \in C_i$
- K-NN cometerá un error si...



El error para k-NN



- Suponga que $x \in C_i$
- K-NN cometerá un error si hay $\frac{k-1}{2}$ o menos patrones en R_i
- La probabilidad de error es

$$p_{ek} = \frac{1}{2^N} \sum_{\alpha=0}^{\frac{k-1}{2}} C_{\alpha}^N$$

Algunas notas

- Comparando p_{e1} y p_{ek} es posible observar que la probabilidad de error de 1-NN es estrictamente menor que k-NN
- Si la distancia entre patrones de una clase es menor a la distancia a patrones de otras clases
- La probabilidad de error de 1-NN es a lo más dos veces la probabilidad de error de Bayes (la menor posible)

Ejercicio

- Desarrolle en Python el clasificador k-NN
- Dados los patrones en \mathbb{R}^2 :
 - $(0,3), (0,2), (0,1), (0,0), (-1,0), (-2,0) \in C_1$
 - $(1,3), (1,1), (1,0), (0,-1) \in C_2$
- Clasifique el patrón $(1,4)$ utilizando
 - 1-NN
 - 3-NN
 - 5-NN
- Para cada caso grafique los patrones e indique los vecinos más cercanos
- Explique los resultados

Para la otra vez...

- Algoritmos para el agrupamiento de datos



iimas

The End.