Reconocimiento Estadístico de Formas

Técnicas No paramétricas

Sección de Reconocimiento de Formas. Grupo de Visión por Computador.

Contenidos

- Introducción
- Clasificación por Distancia Mínima (CDM)
- Regla del vecino mas próximo (1-NN / NN)
- Regla de los *k* vecinos mas próximos *(k*-NN)
- Regla de la vecindad envolvente

Introducción

Introducción: objetivo del RF

objeto o fenómeno físico nuevo

modelo
computacional
entrenado

clase o categoría

Introducción: clasificación

- $\delta \sim \{D_i(x)\},\ i = 1, \ldots, C$
  - $D_i(x)$, criterio de pertenencia a clase i
- $\delta(x) = \omega_i \Leftrightarrow D_i(x) > D_j(x)\ \nabla j \neq i\ i, j = 1, \ldots, C$
- ejemplos:
  - C. paramétrica: $D_i(x) = p(\omega_i \mid x)$
  - C. **no** paramétrica: $D_i(x) = -d\left(x, s_x^i\right),\ s_x^i$ el *NN* de x en ω_i

Introducción: base del problema

Muchas veces no se conocen las características estadísticas del problema. En concreto, no se conoce la distribución que siguen los datos.

Solución

Estimar las funciones de densidad

Estimar las funciones discriminantes

Introducción: c. no paramétrica

dados:

- $T = \{(x, z)\}$, conjunto de entrenamiento
  - $x \in E, \quad z \in \Omega = \{1, 2, \ldots, C\}$
- $d : E\text{x}E \rightarrow \Re$ (metrica entre muestras sobre E)
- x', de clase z' desconocida

regla de decisión $\delta(x') = z'$:

- δ se define a partir de d

Ejemplos de métricas (e. vectoriales)

- Distancia *Euclídea*.

$$d_E(x_i, x_j) = \sqrt{\sum_{k=1}^{d} (x_{i,k} - x_{j,k})^2}$$

- Distancia de *Manhattan* o "*city block*":

$$d_M(x_i, x_j) = \sum_{k=1}^{d} \left| x_{i,k} - x_{j,k} \right|$$

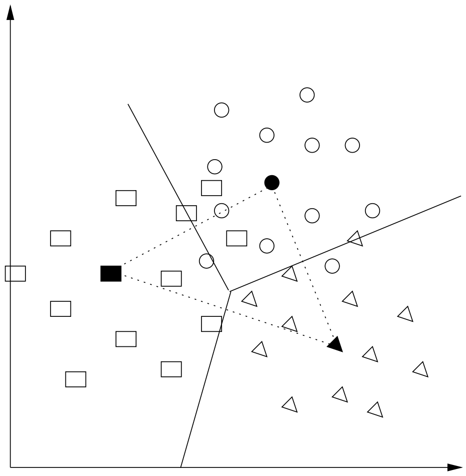
CDM
Clasificación por Distancia Mínima

Clasificación por Distancia Mínima (CDM)

- sea p_i el prototipo representante de la clase i
- ejemplos:
  $$p_i = \arg\min_{p \in T_i} \sum_{x \in T_i} d(p, x) \qquad p_i = \frac{1}{n_i} \sum_{x \in T_i} x$$
- siendo $T_i = \{(x, z) \in T \mid z = i\}$, $n_i = |T_i|$
- regla de clasificación por distancia mínima:
  $$\delta_{DM}(x) = i \Leftrightarrow d(x, p_i) < d(x, p_j) \;\; \forall j \neq i, \;\; i, j \in \Omega$$

CDM: interpretación

- fronteras de decisión $\mathcal{F}$ inducidas CDM:
  - $\mathcal{F} = \{x \mid d(x, p_i) = d(x, p_j) \leq d(x, p_k)\ \forall k \neq j \neq i,\ i, j, k \in \Omega\}$



CDM: descripción de $\mathcal{F}$

- frontera de decisión lineal a intervalos
- frontera entre 2 clases *i, j* = mediatriz de segmento que une p_i, p_j
- frontera entre 2 prototipos $p_{i,}$, p_j
  - limita región de influencia de cada prototipo
  - separa regiones de clasificación de cada clase
- **Diagrama de Voronoi** (VD)/**Polítopos de Voronoi** (PV):
  - estructura definida por $\mathcal{F}$
  - un polítopo por cada p_i delimitado por sus mediatrices
  - $PV(p_i) = \{x \in E : d(x, p_i) \leq d(x, p_j),\ i \neq j,\ i, j \in \Omega\}$

1-NN / NN
Regla del Vecino más Próximo

1-NN: definición

dados:

- $T = \{(x_i, z_i)\}, x_i \in E, \; z_i \in \Omega = \{1, 2, \ldots, C\}$
- $d : E\text{x}E \rightarrow \Re$ (metrica entre muestras sobre E)
- x, de clase z desconocida

regla de decisión *1-NN*:

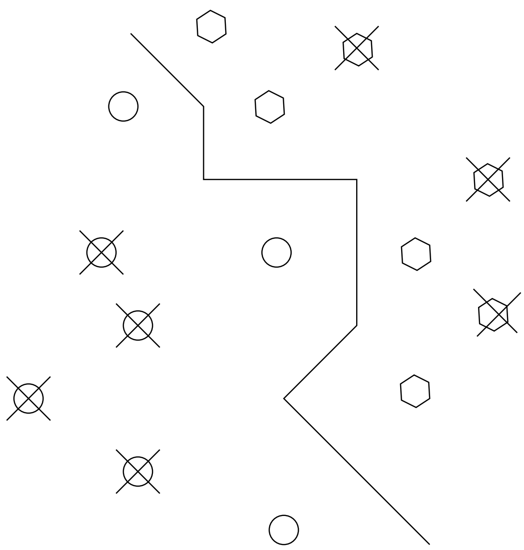
- $z = z_j, \; j = \text{arg min}_{\forall i} \; d(x, x_i)$

1-NN: comentarios

- 1-NN es caso general de CDM
- ventaja respecto a CDM:
  - x se clasifica a partir de todo $(x_i, z_i) \in T$
- su efectividad teórica depende de:
  - número suficientemente grande de prototipos
  - alta densidad local
  - prototipos correctamente etiquetados

1-NN: fronteras de decisión

Frontera lineal a intervalos

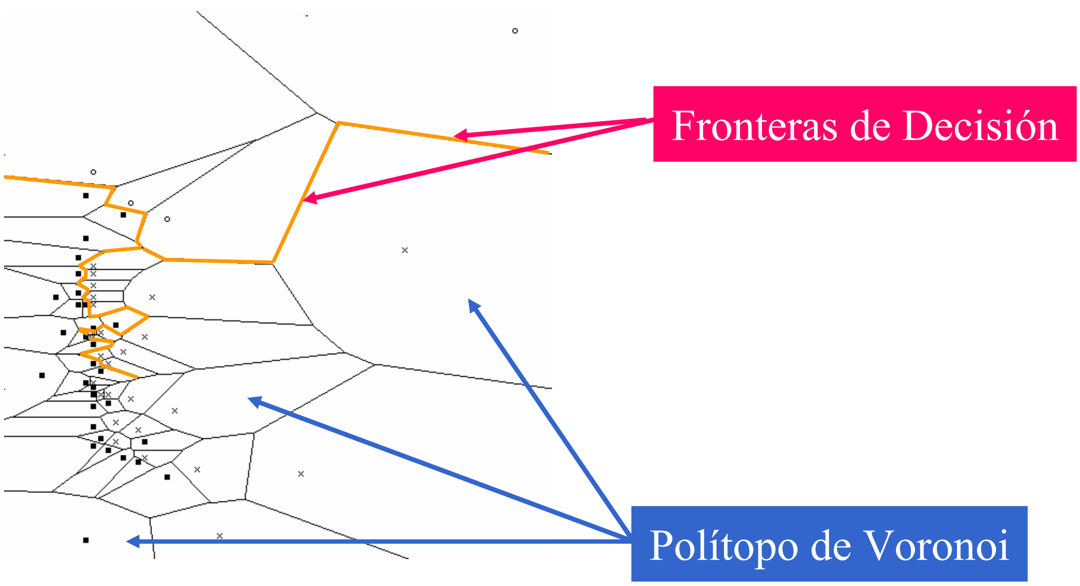


1-NN:

- En este caso, tendremos un polítopo de Voronoi, $P_{j,i}$ para cada prototipo (x_i, j).
- Por tanto, la zona de Voronoi asociada a cada clase vendrá dada por:

$$ZV(P_j) = \bigcap_{i=1}^{n_j} P_{j,i}$$

Diagrama de Voronoi



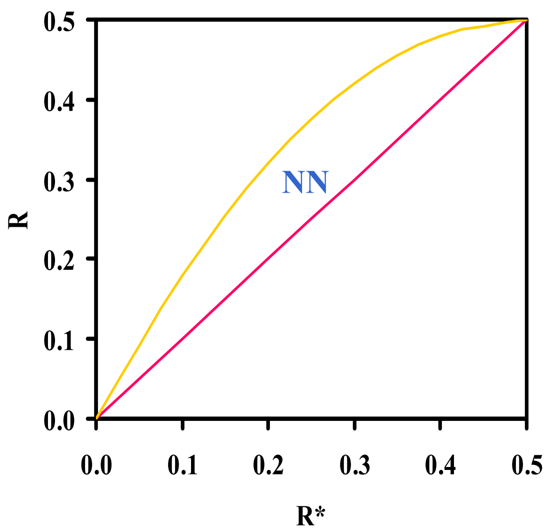
Regla del Vecino más Próximo: Análisis Asintótico

- Análisis de efectividad de la regla NN [Cover, 1967] basado en la hipótesis de disponer de un conjunto de talla ilimitada: $n \to \infty$.
- Puede demostrarse la siguiente relación entre las cotas del riesgo medio asociado a la regla NN (2 clases) :

$$R^* \leq R \leq 2R^*(1 - R^*)$$

donde R^* es la esperanza del riesgo condicional asintótico para la regla de *Bayes* y R para la regla NN.

Regla del Vecino más Próximo: Análisis Asintótico



- Al menos la mitad de la información sobre la pertenencia de un punto a una cierta clase está en su vecino más próximo.
- La cota inferior ($R = R^*$) se obtiene en los dos casos extremos: absoluta certeza ($R^* = 0$) y absoluta incertidumbre ($R^* = 0'5$).

k-NN
Regla de los *k* vecinos más próximos

Regla de los k-Vecinos más Próximos (k-NN)

- Extensión de la regla NN.
- **Objetivo**: usar más información de densidad que NN.
- **Estrategia**: la clase que se asignará a una muestra *x* corresponderá a la clase más votada entre sus *k* vecinos más próximos dentro del conjunto de entrenamiento *T*.
- En general, se tomará un *k* impar para reducir la probabilidad de empates (lo evita con 2 clases).

Regla k-NN

- Sea $T_i = \{(x, z) / z = i\}, i \in \Omega$
- Se puede definir la vecindad $V_k(x)$ de una muestra x como el conjunto de prototipos que cumple:

$$\begin{cases} V_k(x) \subseteq T \\ |V_k(x)| = k \\ \forall p \in V_k(x), \quad q \in T - V_k(x) \Rightarrow d(p,x) \leq d(q,x) \end{cases}$$

Regla k-NN

- Si ahora definimos la distancia entre un punto x y un conjunto de prototipos T_i como

$$d_k(x, T_i) = k - | V_k(x) \cap T_i |$$

entonces podemos definir la regla k-NN como:

$$\delta_{k\text{-}NN}(x) = i \Leftrightarrow d(x, T_i) = \min d_k(x, T_j) \qquad j = 1, ..., c$$

Regla k-NN con Rechazo

- Posibilidad de no clasificar aquellas muestras para las que no se obtiene una cierta garantía de clasificación correcta.
- **Objetivo**: aumentar la efectividad de la regla *k*-NN, descartando la clasificación de las muestras "dudosas", es decir, la de aquellas que se encuentran próximas a las fronteras de decisión.

Regla k-NN con Rechazo

- Se añade una clase artificial *0* que agrupe a las muestras rechazadas. Por tanto, el problema pasa a tener $(c + 1)$ clases.
- En general, se fija un umbral: si la votación no da lugar a una *mayoría cualificada* en alguna de las C clases, la muestra es asignada a *0*.

Regla k-NN con Rechazo I

- Sea l entero positivo, tal que $\lceil k/2 \rceil < 1 \leq k$, un valor umbral para la mayoría en la votación de los k vecinos más próximos. La regla (k, l)-NN [Hellman, 1970] se define como:

$$\delta_{(k,l)-NN}(x) = \begin{cases} i \quad si \quad \left| V_k(x) \cap T_i \right| \geq l \quad i = 1, \ldots, C \\ 0 \quad en\ otro\ caso \quad (clasificación\ rechazada) \end{cases}$$

Regla k-NN con Rechazo II

- **Generalización**: establece un umbral l_i distinto para cada una de las C clases del problema. En este caso, la regla (k, l_i)-NN se define como:

$$\delta_{(k,l_i)-NN}(x) = \begin{cases} i \quad si \quad \left| V_k(x) \cap T_i \right| \geq l_i \quad i = 1, \ldots, C \\ 0 \quad en\ otro\ caso \quad (clasificación\ rechazada) \end{cases}$$

Regla k-NN con Rechazo III

- Establecer una *holgura absoluta* $h \geq 1$: si ninguna de las clases votadas supera al resto en h votos, la muestra será rechazada. La regla (k, h)-NN [Luk, 1986] se define:

$$\delta_{(k,h)-NN}(x) = \begin{cases} i \quad si \quad (k_i - k_j) \geq h \quad \forall j \neq i \quad i, j = 1, \ldots, C \\ \qquad\qquad\qquad\qquad\qquad\qquad \sum_{m=1}^{C} k_m \leq k \\ 0 \quad en\ otro\ caso \quad (clasificación\ rechazada) \end{cases}$$

donde k_i, k_j expresan el número de votos para las clases i, j entre los k vecinos de x.

Regla k-NN con Rechazo III

- **Problema** de la holgura absoluta: se requiere k grande pues, de los contrario, pocas muestras superarían dicho umbral.
- **Solución**: definir una "cooperación" entre la regla k-NN y la regla (k, h)-NN para definir la regla (k, h, k')-NN:

$$\delta_{(k,h,k')-NN}(x) = \begin{cases} i \quad si \quad (k_i - k_j) \geq h \quad \forall j \neq i \quad i, j = 1, \ldots, C \\ \qquad\qquad\qquad\qquad\qquad\qquad \sum_{m=1}^{C} k_m \leq k \\ \delta_{k'-NN}(x) \quad en\ otro\ caso \end{cases}$$

Regla k-NN por Distancia Media

- Aveces tiene más importancia la distancia a la que están los k vecinos que el propio recuento de votos.
- En este caso, se puede definir la distancia entre un punto x y el conjunto de prototipos de una clase como:

$$d_k(x, T_j) = \frac{1}{\left|V_k(x) \cap T_j\right|} \sum_{\forall p \in V_k(x) \cap T_j} d(x, p)$$

- Esta definición corresponde a la distancia media de los prototipos de la clase j de entre los k prototipos más próximos a x. La regla de decisión será la de k-NN.

Análisis Asintótico para la Regla k-NN

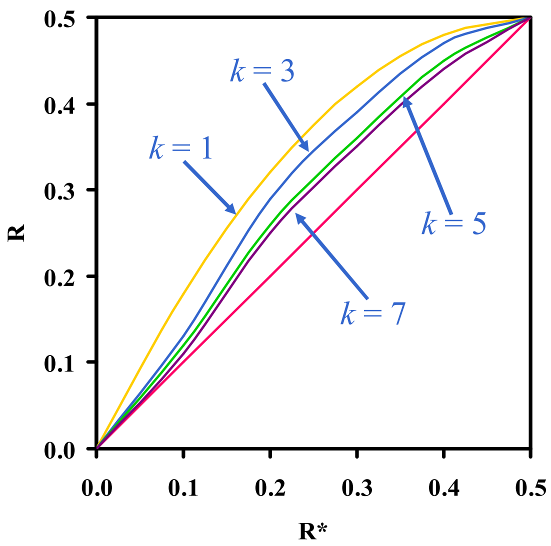
- Consideraremos la siguiente hipótesis:

$$n \to \infty, k \to \infty, k/n \to 0.$$

- Puede demostrarse que:

$$R_{k\text{-NN}} \to R^*$$

Análisis Asintótico para la Regla k-NN



- Conclusión: el error de la regla k-NN decrece a medida que aumentamos el valor de k.
- Por tanto, podemos afirmar que se trata de un **clasificador óptimo** en términos de la regla de *Bayes* (cuando $k \rightarrow \infty$).

Regla de la Vecindad Envolvente

Reglas de Vecindad Envolvente

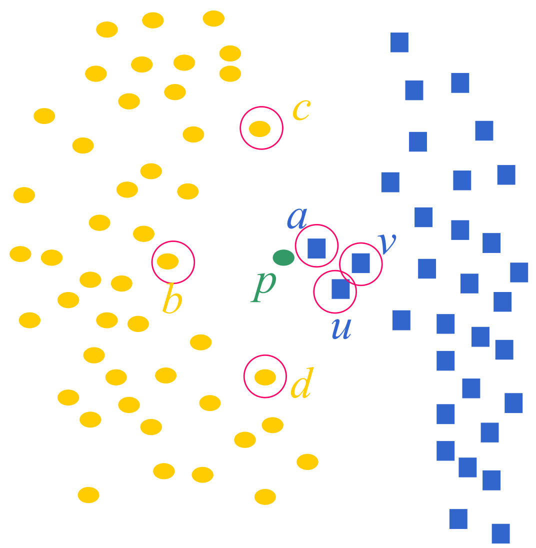
- Se basa en el concepto general de *Vecindad Envolvente* ("*Surrounding Neighbourhood*") [Sánchez, 1997].
- Objetivo: determinar la clase de una muestra a partir de los prototipos que la "rodean" o "envuelven" en el espacio de representación.
- Tratan de resolver las posibles incertidumbres que se producen al clasificar muestras próximas a las fronteras de decisión.

Vecindad Envolvente

- Los vecinos envolventes de una muestra deben satisfacer dos criterios simultáneamente:
  - *<u>Criterio de Distancia</u>*: deben estar tan próximos a la muestra como sea posible.
  - *<u>Criterio de Simetría</u>*: además, deben estar tan homogéneamente distribuidos alrededor de la muestra como sea posible.
- Posibles realizaciones: *Vecindad de Centroide más Próximo*, *Vecindad de Grafo* (*Grafos de Proximidad*).

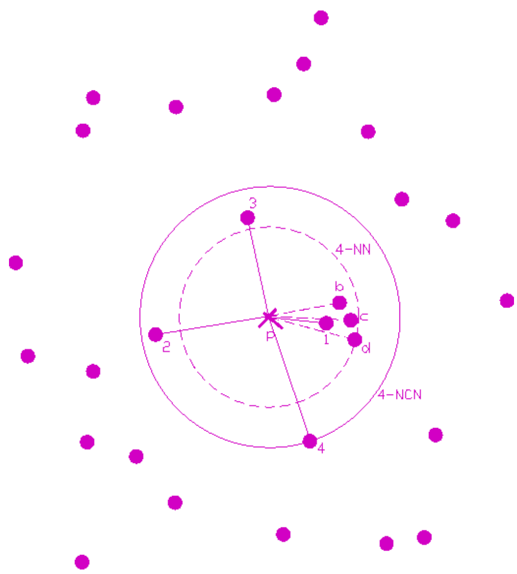
Vecindad de Centroide más Próximo (NCN)

- El primer vecino de un punto *p* será su primer vecino más próximo, q_1.
- El *i*-ésimo vecino de centroide más próximo, q_i, $i > 1$, será el que minimice las distancia entre *p* y el centroide de todos los vecinos, $q_1, \ldots, q_i$, seleccionados antes.



Propiedades del NCN

- El primer NCN coincide siempre con el vecino más próximo.
- El criterio de mínima distancia prevalece sobre simetría, debido a secuencialidad del algoritmo.
- El método es incremental.
- La región definida por la NCN es mayor que la de la vecindad convencional, pero resulta mucho más homogénea (más repartida alrededor de la muestra).



Grafos de Proximidad

- sea
  - $V = \{x\}, x \in R^d, |V| = n$
  - una relación $\wp$: $V \times V$
- grafo de proximidad $G_{\wp}(V, A)$ definido por $\wp$
  - V, conjunto de vértices
  - A es conjunto de aristas $\{(p, q)\}, p \wp q, (p, q) \in V \times V$
  - si $(p, q) \in A$, p y q son vecinos de grafo
  - grafo no dirigido

Ejemplos de Grafos de Proximidad

- Grafo de Gabriel (GG).
- Grafo de Vecindad Relativa (RNG).
- Triangulación de Delaunay (DT).

$$\text{RNG} \subseteq \text{GG} \subseteq \text{DT}$$

Definición de GG

- El conjunto de artistas de un GG, A_{GG}, estará formado por los pares de puntos que satisfagan la siguiente relación $\wp$:

$$(p, q) \in A_{GG} \Leftrightarrow d^2(p, q) \leq d^2(p, r) + d^2(q, r)$$
$$\forall r \in V, \;\; r \neq p, q$$

Representación Geométrica de GG

- La *vecindad de Gabriel* entre dos puntos p y q define una *hiperesfera de influencia* o *hiperesfera diametral* $\Gamma_{p,\,q}$ vacía:

$$\Gamma_{p,q} = B(\frac{p+q}{2}, \frac{d(p,q)}{2})$$

Vecindad de Gabriel

- El conjunto de vecinos de Gabriel ("*Gabriel Neighbours*", GN) para una muestra x respecto a un cierto grafo de Gabriel, $G(V, A)$, estará definido como el conjunto de prototipos tal que:

$$\mathrm{GN}(x) = \{y \in V \,/\, (x, y) \in A\}$$

Definición de RNG

- El conjunto de artistas de un RNG, A_{RNG}, estará formado por los pares de puntos que satisfagan la siguiente relación:

$$(p, q) \in A_{RNG} \Leftrightarrow d(p, q) \leq \max\,[d(p, r), d(q, r)]$$
$$\forall r \in V, \quad r \neq p, q$$

Representación Geométrica de RNG

- La *vecindad relativa* entre dos puntos p y q viene representada por una *hiperluna* $\Lambda_{p,q}$ vacía.

$$\Lambda_{p,q} = B(p, d(p,q)) \cap B(q, d(p,q))$$

Vecindad Relativa

- El conjunto de vecinos relativos ("*Relative Neighbours*", RN) para una muestra x respecto a un cierto grafo de vecindad relativa, $G(V, A)$, estará definido como el conjunto de prototipos tal que:

$$\mathrm{RN}(x) = \{y \in V \,/\, (x, y) \in A\}$$

Reglas de Clasificación Envolventes

- Estrategia: la clase que se asignará a una muestra x corresponderá a la clase más votada entre sus ***k* vecinos de centroide más próximo** o entre sus **vecinos de grafo** dentro del conjunto de entrenamiento T.
- Los correspondientes clasificadores son, en su estrategia, equivalentes a la regla k-NN [Sánchez, 1997]:
  - Regla k-NCN.
  - Regla GN.
  - Regla RN.

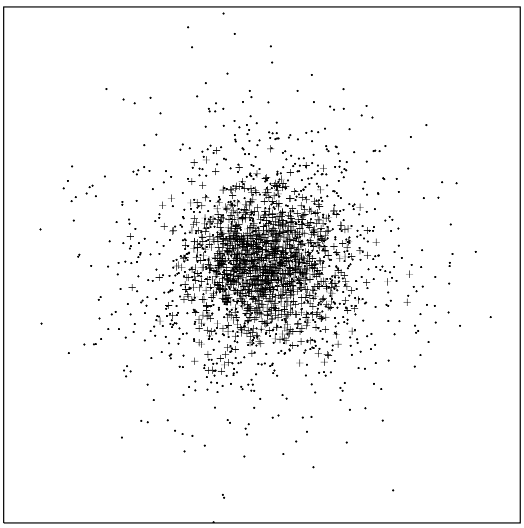
Selección de Prototipos

- Objetivos básicos:
  - Eliminar prototipos erróneamente etiquetados
  - Reducir la talla del conjunto de entrenamiento.

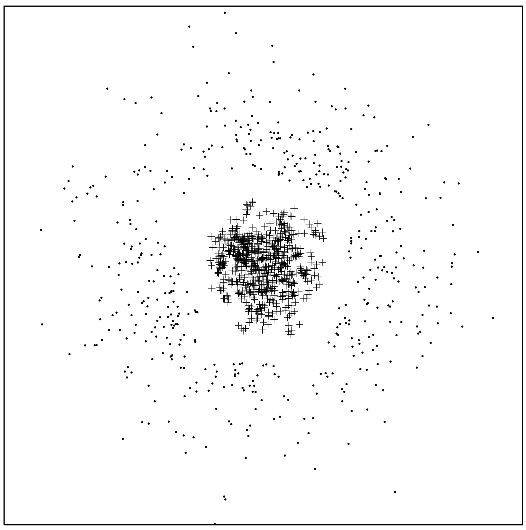
Selección de Prototipos:
familia de técnicas

- Principales técnicas:
  - *Edición* ("*editing*"): descartar prototipos erróneamente etiquetados ("*outliers*") y eliminar solapamientos entre clases distintas. Se tiende a obtener agrupamientos compactos ("*clusters*").
  - *Condensado* ("*condensing*"): reducir drasticamente la talla del conjunto de entrenamiento con el objetivo de reducir el coste computacional de la regla NN, sin perjudicar su tasa de error.

Ejemplo de Edición
2 gaussianas concéntricas solapadas

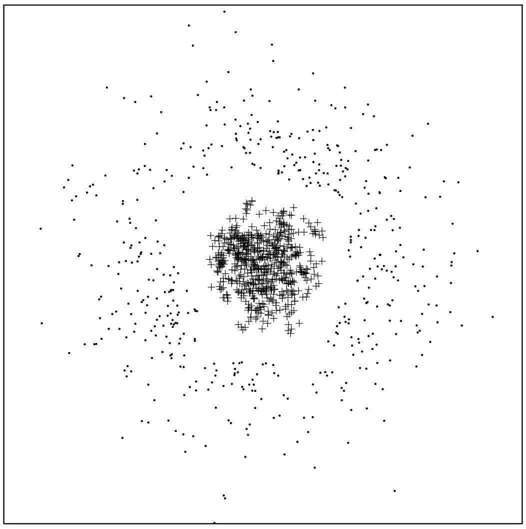


conjunto original con 2500 muestras (65% de acierto con NN sobre test set)

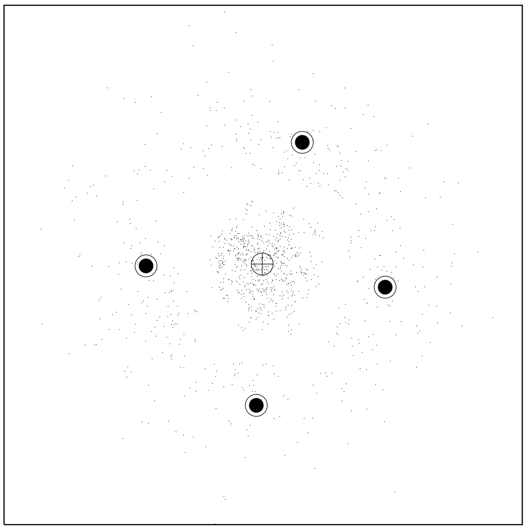


conjunto editado con 1105 muestras (72.2% de acierto con NN sobre test set)

Ejemplo de Condensado (i)
condensado de conjunto editado

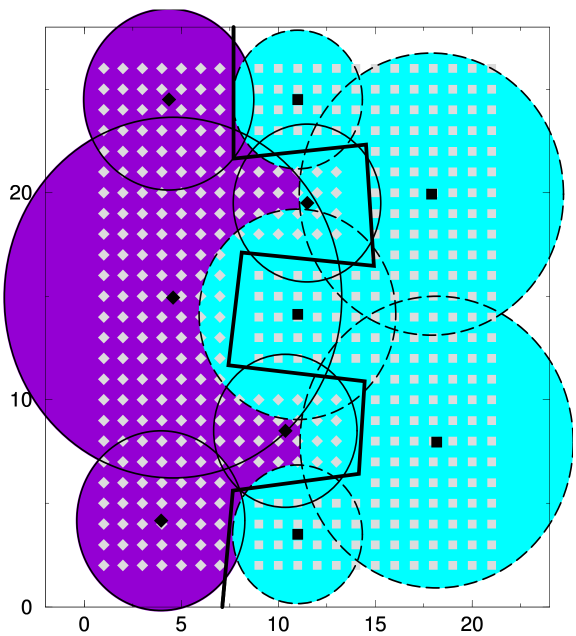


conjunto editado con 1105 muestras
(72.2% de acierto con NN sobre test set)



conjunto condensado con 5 prototipos
(72.9% de acierto con NN sobre test set)

Ejemplo de Condensado (ii)



Problema sintético de 2 clases "entrelazadas" sin solapamiento. El algoritmo de condensado construyó 10 prototipos, la cantidad minima necesaria para definir completamente las fronteras de decisión del clasificador NN.

Algoritmos de Edición

- Objetivos
  - eliminar prototipos mal clasificados y situados en regiones de solapamiento entre clases
  - optimiza capacidad discriminante del NN
  - no persigue grandes reducciones de datos

Algoritmos de Edición

- Edición de Wilson [Wilson, 1972].
- Edición Repetitiva [Tomek, 1976].
- Edición con Re-etiquetado [Koplowitz, 1981].
- Edición con Rechazo [Tomek, 1976].
- Edición por Partición [Devijver, 1980].
- Multiedit [Devijver, 1980].
- Edición mediante Algoritmo Genético [Kuncheva, 1995].
- Edición por Grafos de Proximidad [Sánchez, 1997].
- Edición por NCN [Sánchez, 1997].

Procedimiento General de Edición (i)

Regla de Clasificación

NN
k-NN
(k, l)-NN
k-NCN
GN
RN
Otras ...

+

Estimador del Error

Leaving-one-out
Partición
Validación Cruzada

+

Criterio de Finalización

Iterar
No iterar

Edición de Wilson

<u>ALGORITMO</u> Edición de Wilson (*T*, *k*)

- Inicialización: $S \leftarrow T$.
- Para cada prototipo $x_i \in T$:
  - Buscar los k vecinos más próximos a x_i en $T - \{x_i\}$.
  - Si $\delta_{k\text{-NN}}(x_i) \neq \theta_i$, hacer $S \leftarrow S - \{x_i\}$.

- Estimación mediante "*leaving-one-out*".
- Implementación sencilla.
- Coste computacional $O(n^2)$.

Edición Repetitiva

<u>ALGORITMO</u> Edición Repetitiva (*T*, *k*)

- Inicialización: $S \leftarrow \varnothing$.
- Mientras $|T| \neq |S|$:
  - Pasar el contenido actual de T a S: $S \leftarrow T$.
  - $T \leftarrow$ Edición de Wilson (S, k).

- Repetir la Edición de Wilson hasta que no se produzca ninguna eliminación de prototipos.
- En general, no mejora el resultado de Wilson.

Edición con Re-etiquetado

<u>ALGORITMO</u> Edición con Re-etiquetado (*T*, *k*, *l*)

- Inicialización: $S \leftarrow T$.
- Para cada prototipo $x_i \in T$:
  - Buscar los k vecinos más próximos a x_i en $T - \{x_i\}$.
  - Si $\delta_{(k,\, l)\text{-NN}}(x_i) \neq \theta_i \neq \theta_0$, asignar a x_i la etiqueta de la clase más votada por sus k vecinos más próximos.
  - Si $\delta_{(k,\, l)\text{-NN}}(x_i) \neq \theta_i = \theta_0$, hacer $S \leftarrow S - \{x_i\}$.

- Objetivo: obtener nubes de puntos en cada clase sin solapamientos.

Edición por Partición

<u>ALGORITMO</u> Edición por Partición (T, k, m)

- Hacer una partición aleatoria de T en m bloques: $T_1, \ldots, T_m$.
- Para cada bloque T_j ($j = 1, \ldots, m$):
  - Para cada prototipo $x_i \in T_j$:
    - Buscar los k vecinos más próximos a x_i en $T_{((j+1) \bmod m)}$.
    - Si $\delta_{k\text{-NN}}(x_i) \neq \theta_i$, hacer $T \leftarrow T - \{x_i\}$.

- Estimación mediante partición: cada bloque j se estima a partir del bloque ((j + 1) módulo m).
- Coste computacional $O(n^2/m)$.

Multiedit

- Consiste en iterar la Edición por Partición y utilizar la regla NN (k = 1).
- Comportamiento asintóticamente óptimo, al igual que la versión no repetitiva.

Multiedit

ALGORITMO Multiedit (T, m, f)

- Inicialización: $t \leftarrow 0$.
- Repetir hasta que en las últimas t iteraciones ($t > f$) no se produzca ninguna eliminación de prototipos de T:
  - $S \leftarrow$ Edición por Partición sobre T utilizando NN.
  - Si no se han producido eliminaciones en el paso anterior ($|T| = |S|$), hacer $t \leftarrow t + 1$ e ir al Paso 2.
  - $T \leftarrow S$ y $t \leftarrow 0$.

Resumen Algoritmos Edición

	Regla de Decisión	Método de Estimación	Criterio de Finalización
E. Wilson	k-NN	Leaving-one-out	No iterar
E. Repetitiva	k-NN	Leaving-one-out	Iterar
Re-etiquetado	(k, l)-NN	Leaving-one-out	No iterar
Partición	k-NN	Holdout	No iterar
Multiedit	NN	Holdout	Iterar
E. k-NCN	k-NCN	Leaving-one-out	No iterar
E. Grafos	GN / RN	Leaving-one-out	No iterar

Algoritmos de Condensado

- Objetivo:
  - reducción drástica del numero de prototipos (reduciendo así el esfuerzo computacional)
  - mantener capacidad discriminante del conjunto original
- Clasificación:
  - **por reemplazo**: los prototipos resultantes se construyen a partir de los prototipos originales
  - **por selección**: los prototipos resultantes se seleccionan entre los originales

Algoritmos de Condensado

- Condensado de Hart [Hart, 1968].
- Condensado Reducido [Gates, 1972].
- Condensado Ordenado [Tomek, 1976].
- Condensado por Vecindad Mutua [Gowda, 1979].
- Condensado de Voronoi [Toussaint, 1979].
- Condensado por Grafos de Proximidad [Toussaint, 1985].
- Condensado de Chen [Chen, 1996].
- MCS ("*Minimal Consistent Subset*") [Dasarathy, 1994].
- Condensado Adaptativo (LVQ) [Kohonen, 1990].

Consistencia

- ***Consistencia de conjuntos***: un conjunto de prototipos S es *consistente* respecto a otro conjunto T, si una regla de decisión que use S como conjunto de entrenamiento clasifica T sin errores.

- ***Consistencia de las fronteras de decisión***: un conjunto de prototipos T y un subconjunto de éste serán *consistentes* cuando ambos presentan las mismas fronteras de decisión.

Condensado de Hart (*selección*)

ALGORITMO Condensado de Hart (T) [consistente]

- Inicialización: $S \leftarrow \varnothing$.
- Repetir hasta que no se produzcan más eliminaciones o hasta que el conjunto T esté vacío.
  - Para cada prototipo $x_i \in T$:
    - Buscar el vecino más próximo a x_i en S.
    - Si $\delta_{NN}(x_i) \neq \theta_i$, hacer $T \leftarrow T - \{x_i\}$ y $S \leftarrow S + \{x_i\}$.

Condensado de Hart

- Elimina los prototipos que no resultan necesarios para la correcta clasificación del resto de puntos de T.
- Se supone que si un punto es erróneamente clasificado, se debe al hecho de encontrarse próximo a las fronteras y, por tanto, no debe eliminarse.
- En general, con muy pocas iteraciones se obtiene un conjunto consistente $\Rightarrow$ algoritmo rápido.
- En general, la talla del conjunto condensado es mucho más pequeña que la del conjunto original, aunque no se asegura que sea **minimal**.

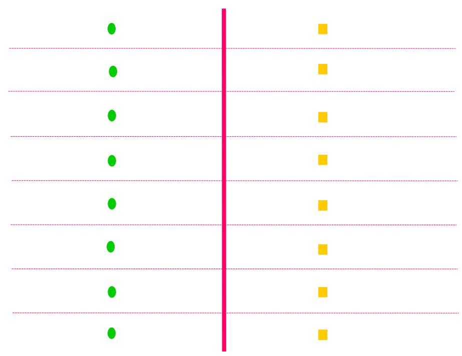
Condensado de Voronoi

ALGORITMO Condensado de Voronoi (T)

- Calcular el VD asociado a T.
- Para cada prototipo $x_i \in T$:
  - Para cada prototipo x_j vecino de Voronoi de x_i:
    - Si $\theta_i \neq \theta_j$, marcar x_i e ir al Paso 2.
- Eliminar todos los prototipos **no marcados**.

Condensado de Voronoi

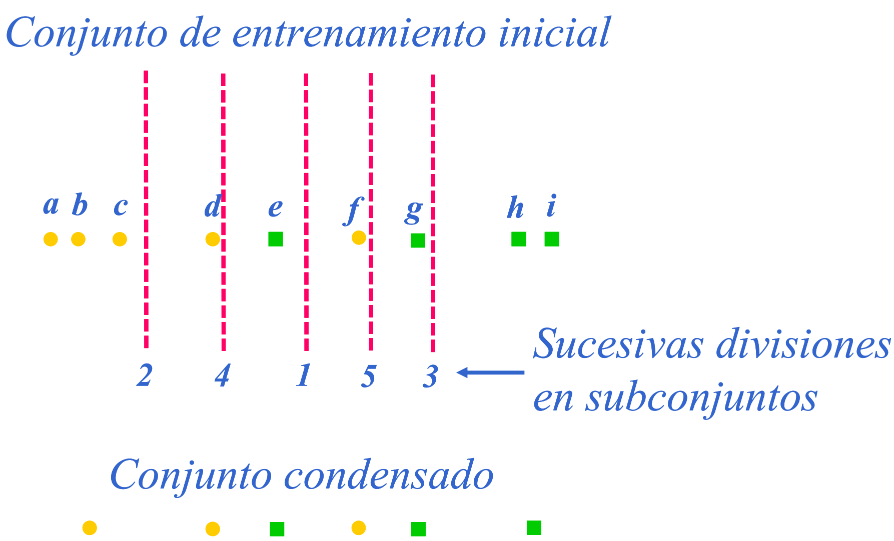
- Obtiene un conjunto reducido y consistente respecto a las fronteras de decisión.
- Sin embargo, no siempre es el conjunto minimal.



Condensado de Chen (*reemplazo*)

- Permite fijar de antemano el número de prototipos a seleccionar.
- *Diámetro de un conjunto*: distancia entre sus dos prototipos más alejados.
- Estrategia: dividir el conjunto original en sucesivos subconjuntos a partir de su diámetro y sustituir cada uno de los subconjuntos por su centro de gravedad, asignándole la etiqueta de clase de la mayoría de sus prototipos.

Condensado de Chen *(reemplazo)*



Condensado Adaptativo *(reemplazo)*

- construyen prototipos que aproximan de manera óptima las distribuciones de probabilidad de cada clase.
- genera nuevos prototipos.
- ajuste iterativo (adaptación) de la posición de los prototipos a partir de un conjunto de entrenamiento
- Ejemplos: métodos LVQ (*Learning Vector Quantization*).

Método LVQ

- Buscar muestra original x mas próxima a prototipo m_i.
- En cada iteración, recompensar o penalizar m_i (*con factor de corrección*, $0 < \alpha < 1$) según coincida o no su clase con la de x.

$$m_i(t+1) = m_i(t) \pm \alpha(t)(x(t) - m_i(t))$$

  - Recompensa: acercar m_i a x.
  - Penalización: alejar m_i de x.