Modelos No Par

Sistemas Inteligentes I

Tema 9. Modelos No Paramétricos

José A. Montenegro Montes
monte@lcc.uma.es

Resumen

- Introducción
- Modelo vecino más próximo
- Regresión no paramétrica
- Conclusiones
- Bibliografía

Introducción

Motivación

- El clima mundial se estudia mediante imágenes obtenidas por satélites
- Estas imágenes tienen defectos:
  - Hay píxeles ausentes si un satélite no pudo obtener datos un cierto día
  - Las limitaciones de los sensores producen ruido

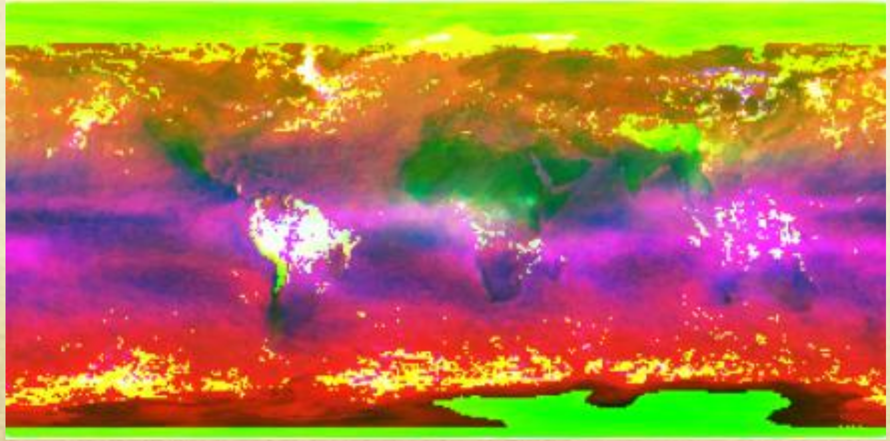


Imagen de clima mundial del *Earth Observatory* (NASA). Rojo = fracción de nubes, verde = monóxido de carbono, azul = vapor de agua.

Motivación

- Las personas pueden completar estas imágenes correctamente, pero no muy rápido, o bien completarlas rápido, pero no muy correctamente
- Por ejemplo, si una persona intentase rellenar las imágenes de concentración de ozono obtenidas en un día de la forma más precisa posible, tardaría 500 años

Introducción

- Redes neuronales utiliza información de entrenamiento para establecer un conjunto fijo de parámetros ***w***.
- La información de entrenamiento es descartadas ya que viene representada por el parámetro ***w***.
- El modelo de aprendizaje que resume la información con un conjunto de parámetros de tamaño fijo es denominado modelo paramétrico.
  - Independiente del número de ejemplos de entrenamiento.

Introducción

- Un **modelo no paramétrico** es aquel que no es representado por un conjunto limitado de parámetros.
  - Cada hipótesis generada mantenemos todos los datos y serán usadas para predecir el siguiente ejemplo.
  - Tal hipótesis es no paramétrico ya que el número de parámetros no está limitado y crece con el número de ejemplos.
  - Este ejemplo es denominado **aprendizaje basado en el ejemplo** o **aprendizaje basado en memoria**.
    - Ejemplo más simple es una tabla de búsqueda, con toda la información de entrenamiento. Cuando preguntamos por ***h(x)***, si ***x*** está en la tabla devolvemos el correspondiente ***y***.

Modelos No Par

Modelo vecino más próximo

Modelo vecino más próximo

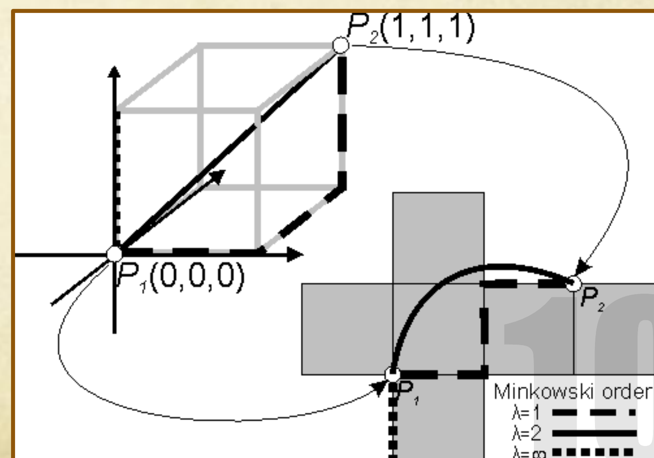
- **Búsqueda de los k vecinos más cercanos**: Podemos mejorar el ejemplo anterior con una pequeña variación, dado un punto de consulta x_q, encontrar los k ejemplos que están más cerca de x_q. **Definimos**: NN(k, x_q);
- Utilización:
  - Encontrar los NN(k, x_q).
    - **Clasificación**: Escoger la mayoría de los casos. Para evitar empates, es aconsejable escoger k un número impar.
    - **Regresión**: Escogemos la media o mediana de los k vecinos, o resolver un problema de regresión lineal de los vecinos.

Distancia Minkowski

- **<u>Distancia Minkowski</u>** entre los datos p y k, donde n es el número de dimensiones (atributos), k el índice de la variable, n total de variables y λ el orden.

$$dist = \sqrt[\lambda]{\sum_{k=1}^{n} (p_k - q_k)^{\lambda}}$$

- $\lambda = 1$ es la distancia Manhattan
- $\lambda = 2$ es la distancia Euclidea
- $\lambda = \infty$ es la distancia Chebyshev
  - Max $p_k - q_k$



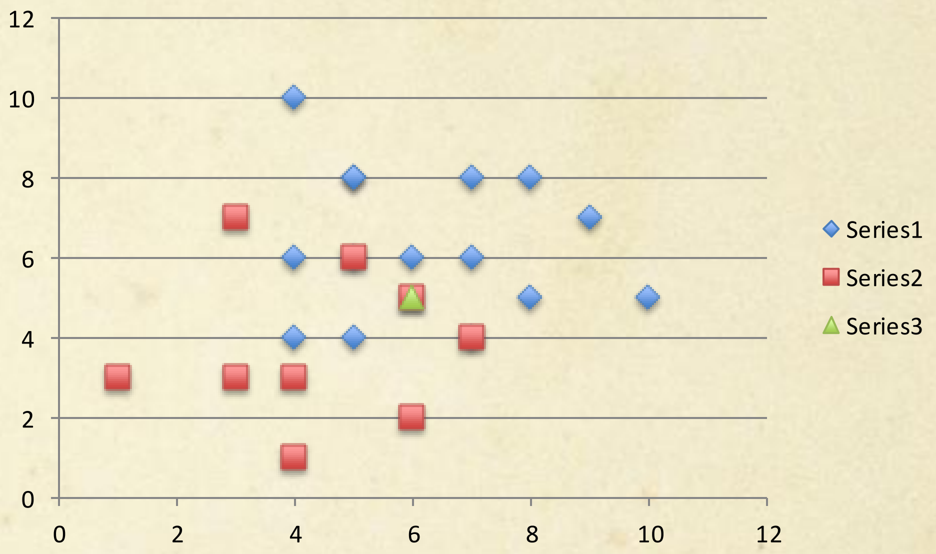
Ejemplo Distancia Minkowski

	Variable1	Variable2	Variable3
Caso1	1	1	1
Caso2	1	1	0
Caso3	2	2	2
Caso4	10	10	10
Caso5	11	11	11
Caso6	10	5	0

2	Caso1	Caso2	Caso3	Caso4	Caso5	Caso6
Caso1	0	1,00	1,73	15,59	17,32	9,90
Caso2	1,00	0	2,45	16,19	17,92	9,85
Caso3	1,73	2,45	0	13,86	15,59	8,77
Caso4	15,59	16,19	13,86	0	1,73	11,18
Caso5	17,32	17,92	15,59	1,73	0	12,57
Caso6	9,90	9,85	8,77	11,18	12,57	0

Ejemplo KNN(8)

X1	X2		Distancia	Rank	
6	6	+	1,00	2	
5	4	+	1,41	3	
7	4	+	1,41	4	
5	6	+	1,41	5	
7	6	+	1,41	6	5
8	5	+	2,00		
4	4	+	2,24		
4	6	+	2,24		
5	8	+	3,16		
5	8	+	3,16		
7	8	+	3,16		
9	7	+	3,61		
8	8	+	3,61		
10	5	+	4,00		
4	10	+	5,39		
6	5	-	0,00	1	3
7	4	-	1,41	7	
5	6	-	1,41	8	
4	3	-	2,83		
6	2	-	3,00		
3	3	-	3,61		
3	7	-	3,61		
4	1	-	4,47		
1	3	-	5,39		
6	5				



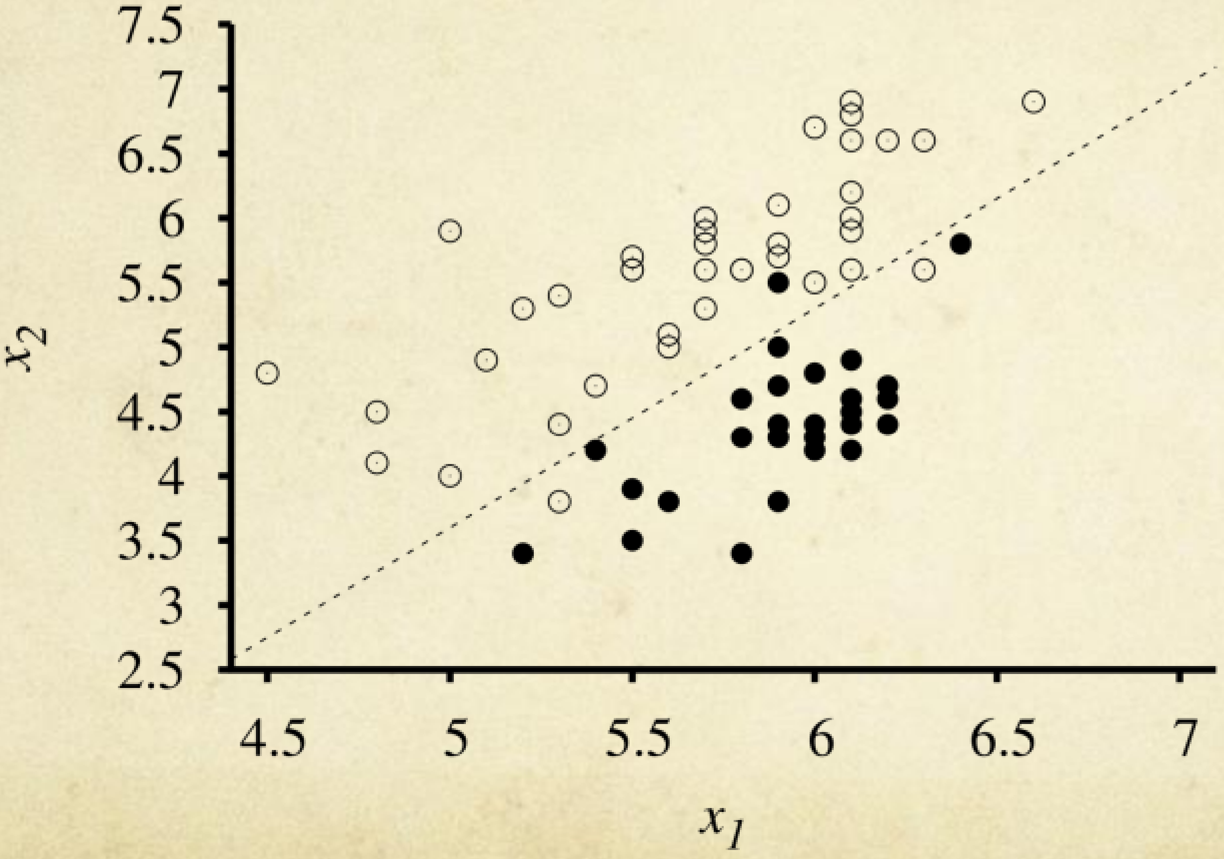
Distancias

- Los vecinos más próximos funcionan bien para espacios de entrada de dimensión **n** pequeña.
- Para espacios de elevada dimensionalidad los vecinos más próximos no tienen buenos resultados, ya que los vecinos están muy alejados.
  - **Maldición de la dimensionalidad.**

Ejemplo: terremotos (I)

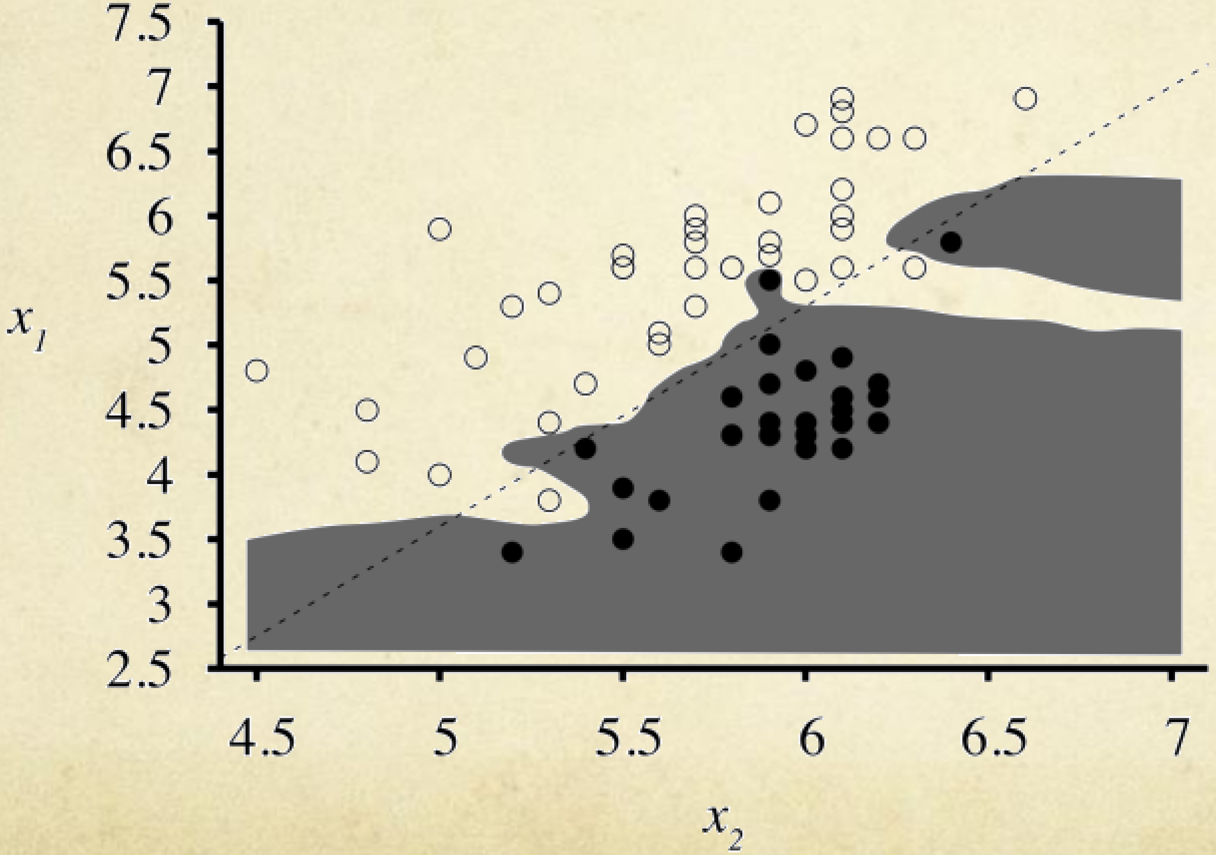
- Tenemos puntos de datos de dos clases: terremotos y explosiones nucleares subterráneas
- Dada la magnitud de la onda interna x_1 y la magnitud de la onda superficial x_2, queremos decidir (clasificar) si un suceso es un terremoto o una explosión
- En las siguientes figuras, los terremotos se marcan con círculos blancos y las explosiones con círculos negros

Ejemplo: terremotos (II)



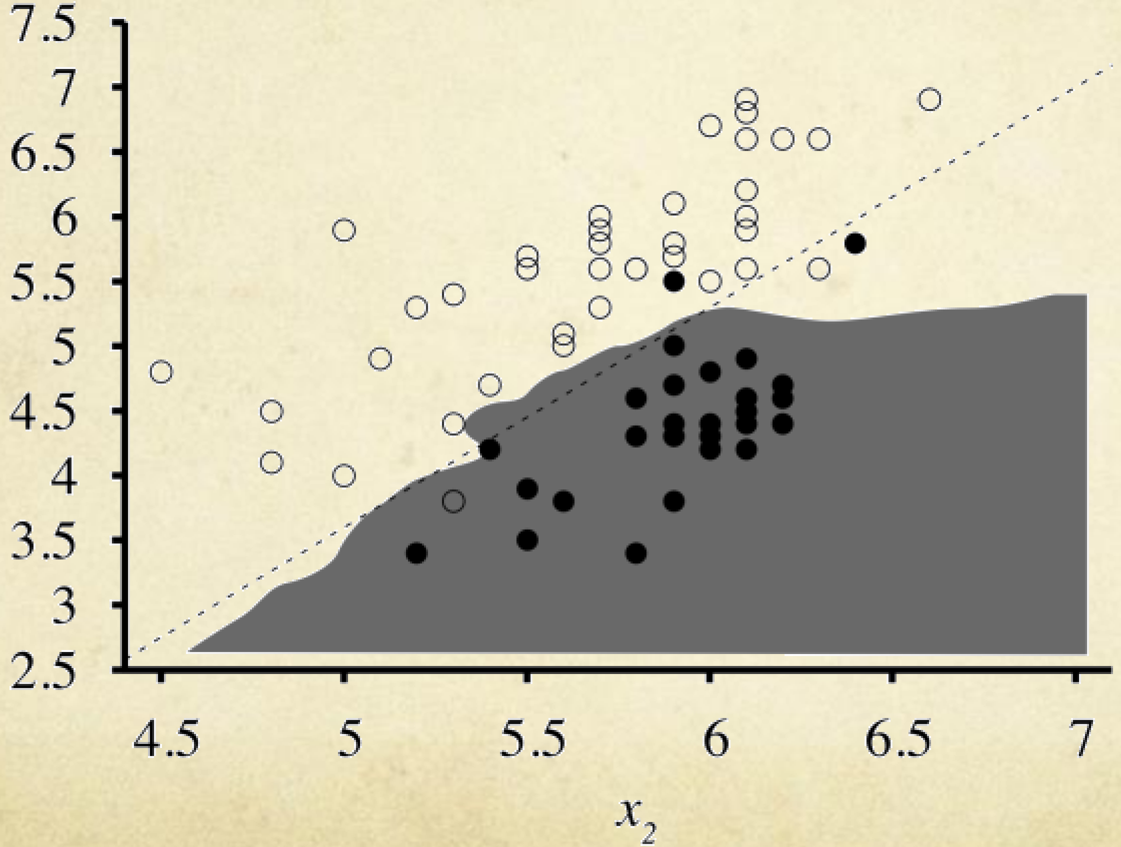
Ninguna línea recta separa las clases: las clases no son linealmente separables

Ejemplo: terremotos (III)



Resultado de *NN* con *k*=1 (se produce un exceso de ajuste)

Ejemplo: terremotos (IV)



Resultado de *NN* con *k*=5 (el exceso de ajuste desaparece)

Regresión no paramétrica

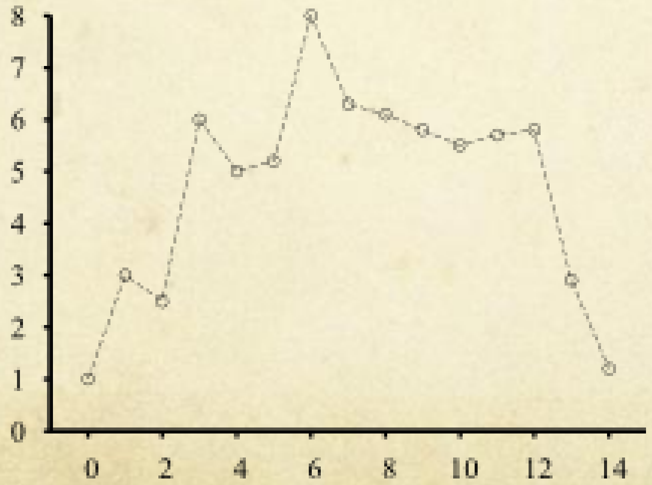
Regresión no paramétrica

- Existen varios modelos para la regresión no paramétrica.

  - **Conecta los puntos**
  - **k vecinos más cercanos**
  - **Regresión con ponderación local**

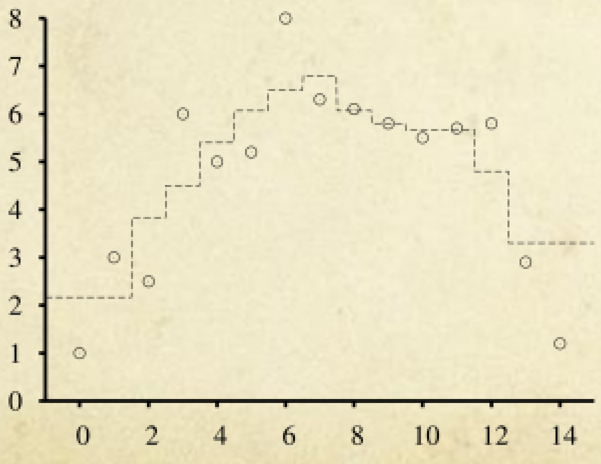
Regresión no paramétrica

- **Conecta los puntos:** Es el modelo más simple.
  - Crea una función h(x) que dado un punto consulta x_q, resuelve el problema de regresión con dos puntos de los ejemplos de entrenamiento, a la izquierda y derecha de x_q.
  - Método utilizado en las hojas de cálculo.

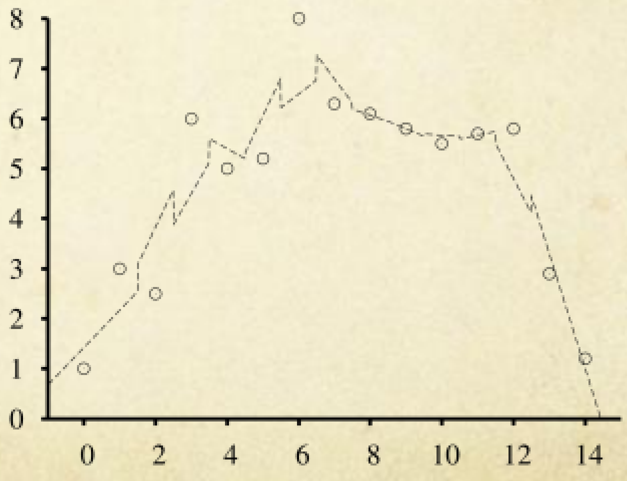


Regresión no paramétrica

- **k vecinos más cercanos:** mejora el anterior (k=3).
  - (a) Utilizamos los k vecinos más cercanos, en vez de utilizar dos puntos del punto x_q. h(x) es la media de los k puntos.
  - (b) Regresión lineal de los k vecinos más cercanos



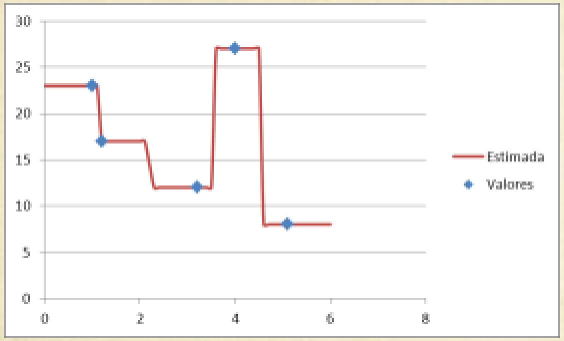
a



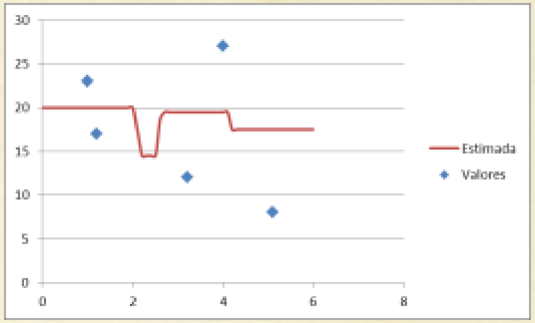
b

Regresión no paramétrica

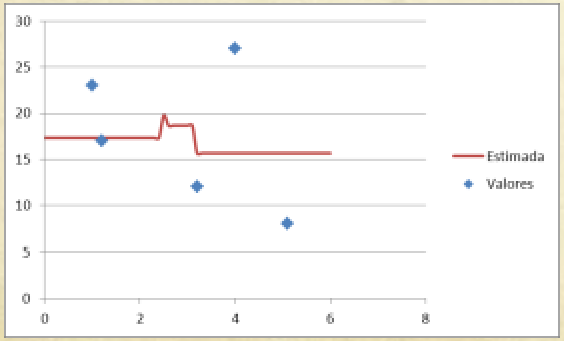
X	1	1,2	3,2	4	5,1
Y	23	17	12	27	8



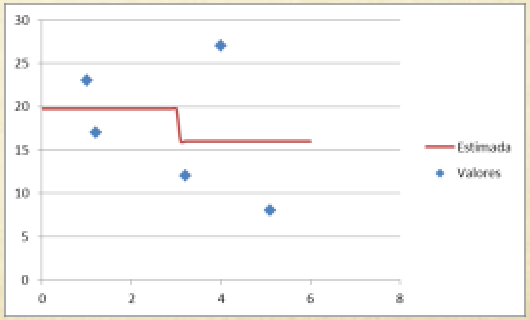
K=1



K=2



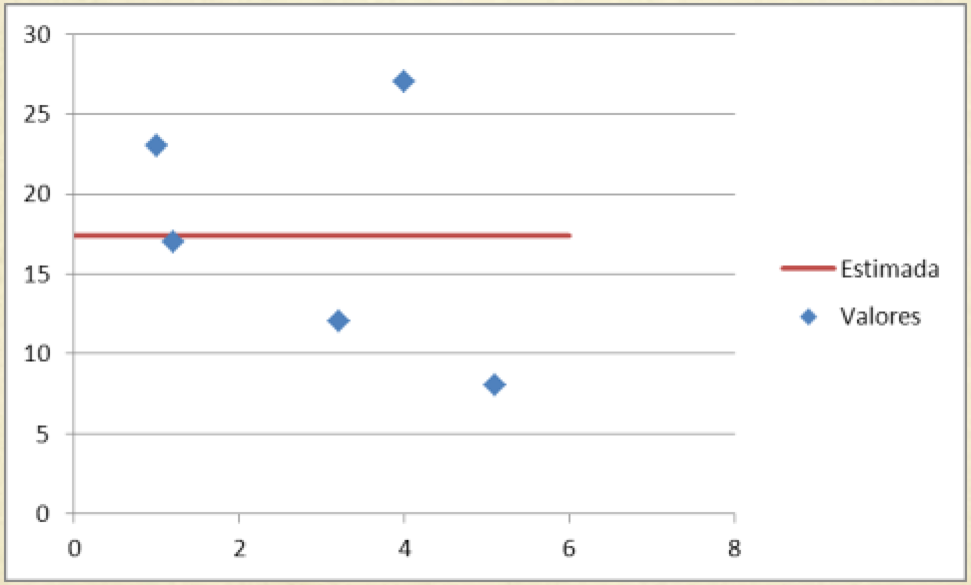
K=3



K=4

Regresión no paramétrica

X	1	1,2	3,2	4	5,1
Y	23	17	12	27	8



K=5

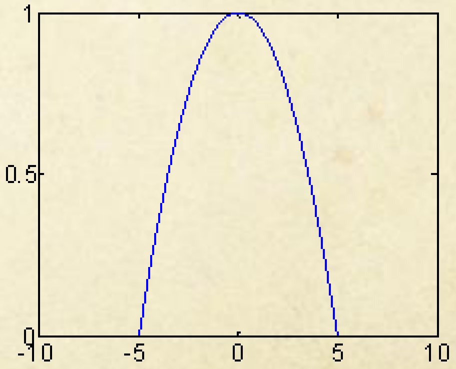
Regresión con ponderación local

- Podemos hacer regresión con los vecinos más próximos calculando la media o bien llevando a cabo una regresión lineal
  - Se capta la tendencia general, pero el resultado es discontinuo
- La regresión con ponderación local nos ofrece las ventajas de los vecinos más próximos, pero sin las discontinuidades
  - El resultado es no lineal
- La idea básica es ponderar más a los puntos más cercanos a x_p y menos a los más lejanos, siendo el decremento gradual.

Regresión con ponderación local

- Decidimos cuanto ponderar mediante una función conocida como ***kernel***.
- Tendremos $K(Distance(\mathbf{x}_j,\mathbf{x}_q))$; K debe ser simétrica alrededor de 0 y tener un máximo en 0

$$K(d) = \max\left(0, 1 - \left(\frac{2|x|}{k}\right)^2\right)$$



El núcleo cuadrático con anchura del núcleo k=10

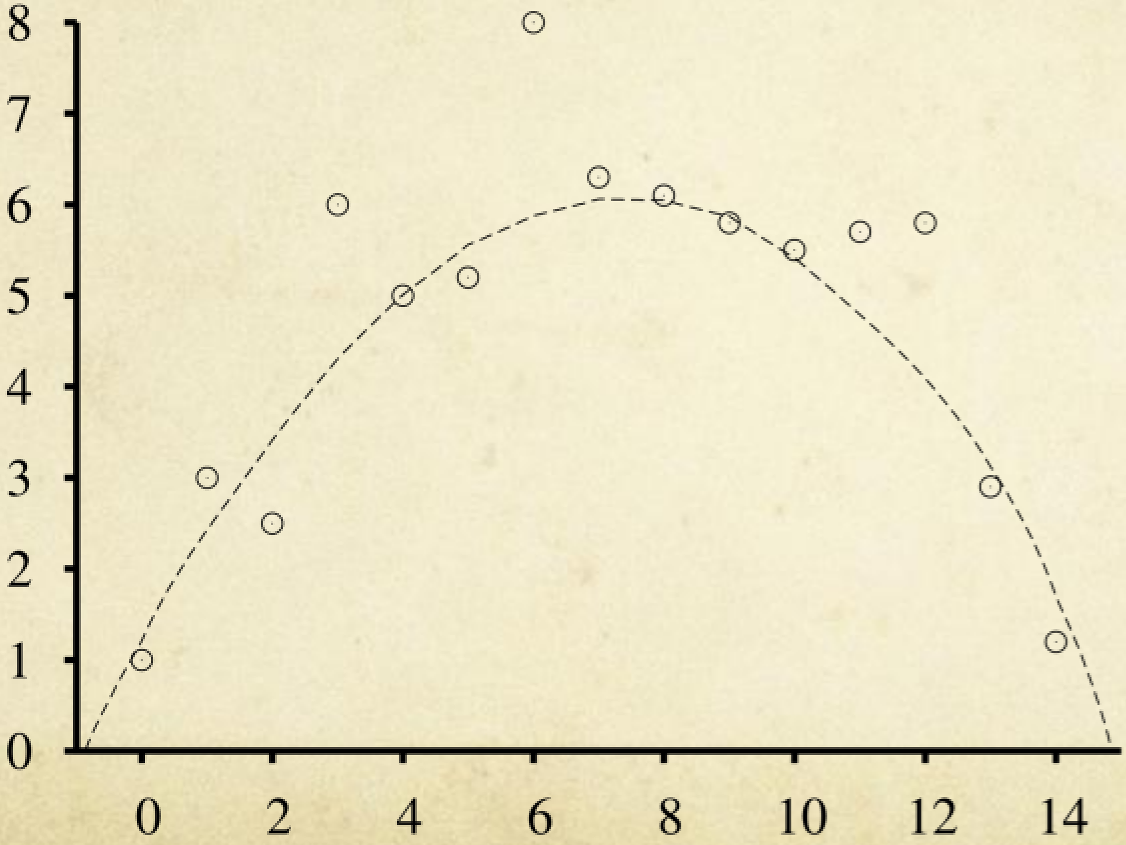
Regresión con ponderación local (I)

- Para un punto de consulta dado $\mathbf{x}_q$, la respuesta es $h(\mathbf{x}_q)=\mathbf{w}^*\mathbf{x}_q$
- El vector de pesos local $\mathbf{w}^*$ se obtiene resolviendo el siguiente problema de regresión ponderada,
  - donde *Distance* es cualquiera de las medidas de distancia consideradas anteriormente:

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} E(\mathbf{w})$$

$$E(\mathbf{w}) = \sum_j K\left(Distance\left(\mathbf{x}_j, \mathbf{x}_q\right)\right)\left(y_j - \mathbf{w}\cdot\mathbf{x}_j\right)^2$$

Ejemplo de regresión con ponderación local



Regresión con ponderación local, empleando un núcleo cuadrático de anchura k=10

Ejemplo de regresión con ponderación local

X	1	1,2	3,2	4	5,1
Y	23	17	12	27	8
weight	1	1	1	1	1

Kernel Gausiano

$$k_\alpha(x, X) = e^{\left(-\frac{(x-X)^2}{2\alpha^2}\right)}$$

Nadaraya-Watson

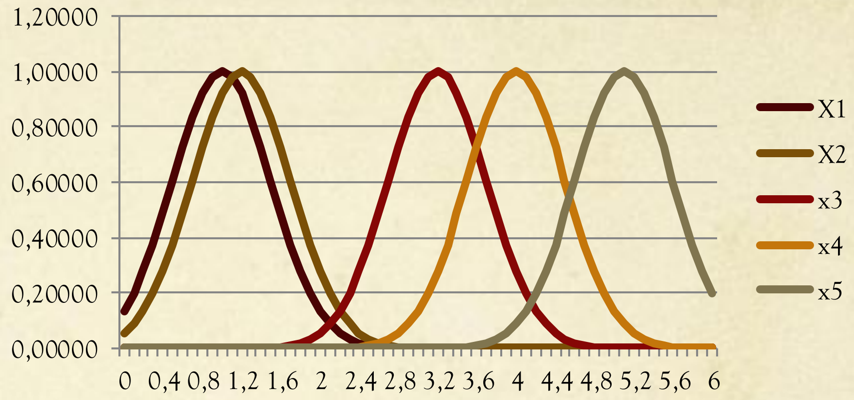
$$\hat{y}_j = f(x, w) = \frac{\sum_{j=1}^{n} w_i K_\alpha(x_j, X_i)}{\sum_{j=1}^{n} K_\alpha(x_j, X_i)}$$

Solver Minimizando SSE Obtengo los pesos

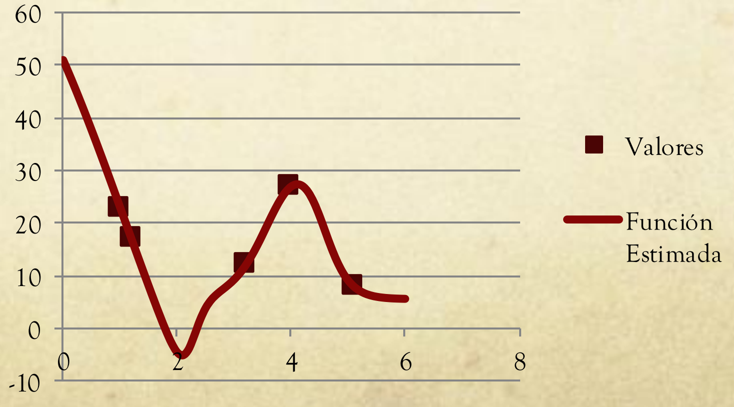
X	1	1,2	3,2	4	5,1
Y	23	17	12	27	8
weight	95,0205109	-55,017975	5,674135	34,830914	5,61586225

Ejemplo de regresión con ponderación local

Kernel Gausiano



X	1	1,2	3,2	4	5,1
Y	23	17	12	27	8
weight	95,0205109	-55,017975	5,674135	34,830914	5,61586225



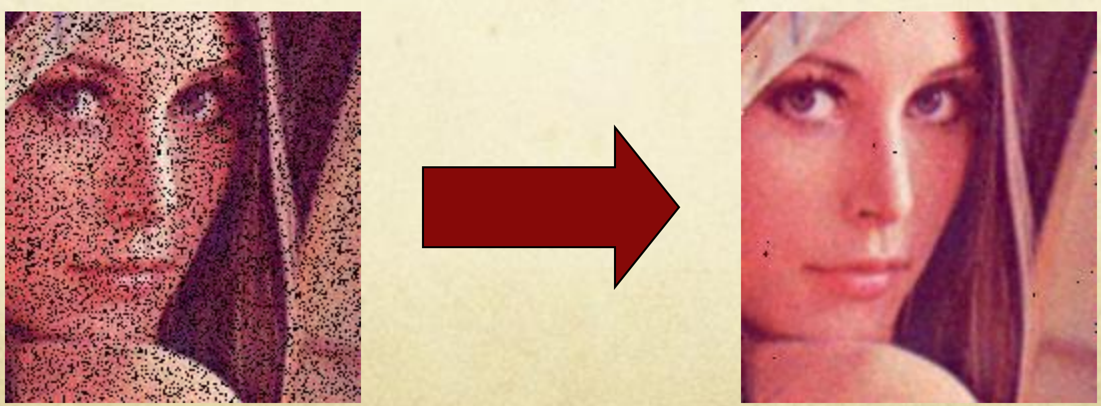
Conclusión

Sumario

- Los modelos no paramétricos usan todos los datos para hacer cada predicción, en lugar de intentar resumir previamente los datos en unos pocos parámetros
- Los vecinos más próximos pueden usarse para la clasificación y para la regresión cuando se emplea una medida de distancia adecuada
- La regresión con ponderación local produce aproximaciones suaves de funciones continuas, y es capaz de manejar datos ausentes y ruido

Epílogo

- Las técnicas actuales de restauración de imágenes procesan las imágenes de satélite obtenidas en un día en cuestión de minutos
- La regresión no paramétrica es una de estas técnicas, como veremos en el laboratorio



Bibliografía

Bibliografía

- AIMA 3 Edición

Sistemas Inteligentes

José A. Montenegro Montes

monte@lcc.uma.es