



TRATAMIENTO DIGITAL DE SEÑALES

Ingeniería de Telecomunicación (4º, 2º c)

Unidad 1ª: Introducción a la decisión máquina

Aníbal R. Figueiras Vidal

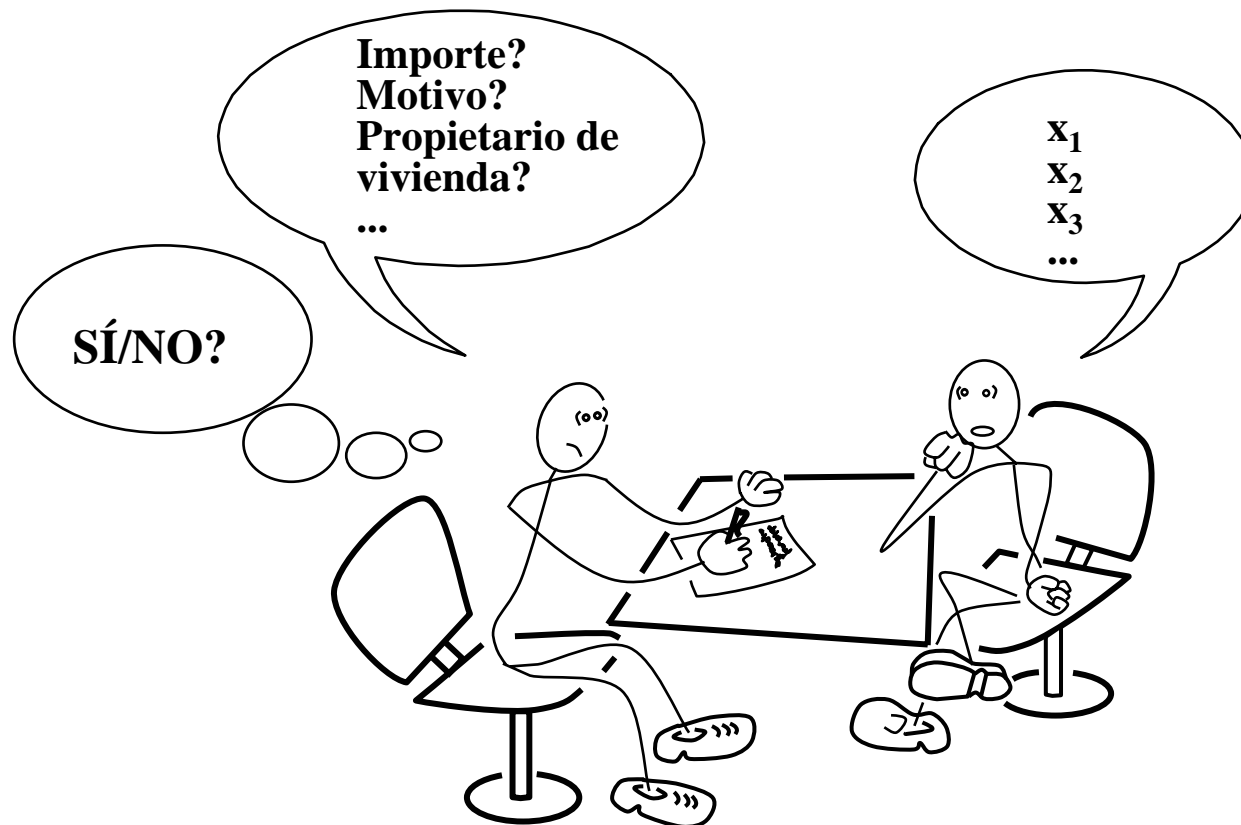
Jesús Cid Sueiro

Ángel Navia Vázquez

Área de Teoría de la Señal y Comunicaciones
Universidad Carlos III de Madrid

Un problema de tratamiento de datos

Concesión de un crédito:





- Quien concede o no el crédito pretende evitar la morosidad... pero no puede determinarla directamente
- Obtiene información recurriendo a otras **variables relacionadas** con ella.
- A partir de las cuales evalúa, implícita o explícitamente, una **función**

$$F_w(x_1, x_2, \dots, x_N) = F_w(\mathbf{x})$$

que, en este caso (sí o no: binario), puede tomar dos valores (típicamente: 0/1; ±1).

- Los **parámetros w** se eligen para conseguir resultados deseables o aceptables: junto con la forma de F, determinan la **capacidad expresiva** (calidad de los resultados alcanzables) de que se dispone.



- Tradicionalmente, en la evaluación de créditos se ha utilizado

$$F_w(\mathbf{x}) = u(\mathbf{w}^T \mathbf{x} - \eta)$$

que equivale a decidir de acuerdo con una comparación (**test de umbral**)

$$\mathbf{w}^T \mathbf{x} \begin{matrix} \text{SÍ} \\ \gtrsim \\ \text{NO} \end{matrix} \eta$$

eligiendo los pesos \mathbf{w} y el umbral η para maximizar el beneficio (utilizando $\mathbf{x}_e^T = [1 \ \mathbf{x}]$, $\mathbf{w}_e^T = [-\eta \ \mathbf{w}]$, $-\eta$ es un parámetro más).

- La función $\mathbf{w}^T \mathbf{x}$ (ó $\mathbf{w}_e^T \mathbf{x}_e$) que permite proceder al compararla con un umbral recibe el nombre de **discriminante** (si hubiese más opciones, los discriminantes se compararían entre ellos)



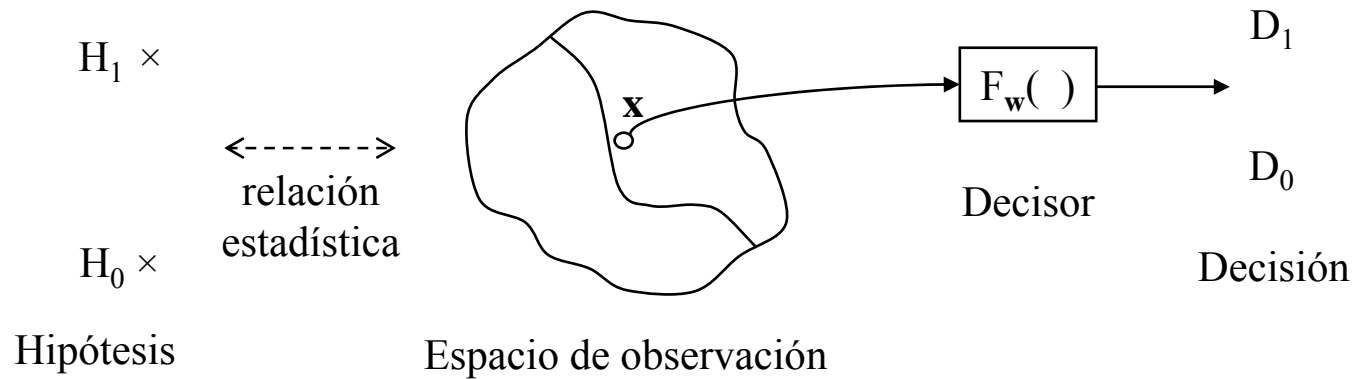
Nomenclatura

- Cada solicitante: caso, instancia o ejemplo; (k)
- Su conjunto de respuestas: observación o dato; $\mathbf{x}^{(k)}$ (muestra)
- Cada componente de las respuestas: variable o rasgo; x_n
- Las posibilidades (mora o no): hipótesis o clases; H_i, C_i



- La función F_w : decisor o clasificador
- El resultado (concesión o denegación): decisión o clasificación; D_j
- El proceso: igualmente, **decisión o clasificación**
(**detección**: presencia o ausencia)
- El número de clases: dimensión del proceso; C
- El número de casos para el diseño: tamaño de la muestra; K
- El número de variables: dimensión del espacio muestral; N

Visión de la decisión (máquina)



La decisión equivale a trazar una frontera en el espacio de observación (habitualmente, las decisiones son exhaustivas y excluyentes).



Nomenclatura complementaria

- Cuando la muestra para el diseño se toma de una (gran) base de datos:
Minería de Datos (“Data Mining”, DM)
- Cuando lo que se pretende es seleccionar ítems de interés de una cierta clase de información (textual, gráfica, de audio, de vídeo, etc.) de una base no estructurada (no indexada, no etiquetada), se habla de **Recuperación de Información** (“Information Retrieval”, IR).
En tal caso, es esencial la selección de **rasgos**.
- La detección o clasificación de rasgos (especialmente de imágenes) se denomina **Reconocimiento de Formas** (“Pattern Recognition”, PR).



Discusión

Q: ¿Conviene utilizar el mayor número de variables accesibles?

Obviamente, no interesa incorporar variables

- *irrelevantes: no relacionadas con las hipótesis;*
- *redundantes: que aporten igual información*

ya que darían lugar a imprecisiones o incertidumbres en el diseño.

Q: ¿Conviene utilizar en el diseño muchas observaciones, o mejor pocas?

*El conjunto de entrenamiento (de diseño del decisor) ha de ser, lógicamente, **representativo** en términos estadísticos: pero no más abundante de lo necesario para ello, para evitar sobrecargas computacionales.*

*También es verdad que hay ejemplos más **críticos** que otros para obtener un buen diseño: en este caso, los que más ayudan a una buena definición de la frontera de decisión.*



Q: ¿Es el objetivo forzoso de un problema de decisión minimizar el número de errores?

*No: en general, es optimizar un **objetivo**; así, en el caso de concesión de créditos, maximizar el beneficio global.*

Q: ¿Debe perseguirse la estricta optimización del objetivo sobre el conjunto de entrenamiento?

*No: el propósito del diseño es que resulte útil para casos no contemplados previamente: esta capacidad de **generalización** se contrapone con un diseño muy “a la medida”.*



Q: ¿Cómo determinaría los parámetros w del clasificador?

*Mediante la aplicación de un algoritmo de **búsqueda** (local o global) para optimizar el objetivo (¡y mantener la generalización!). En ocasiones, resultan aplicables soluciones bloque.*

Q: ¿Cómo elegiría la forma F de la función de decisión?

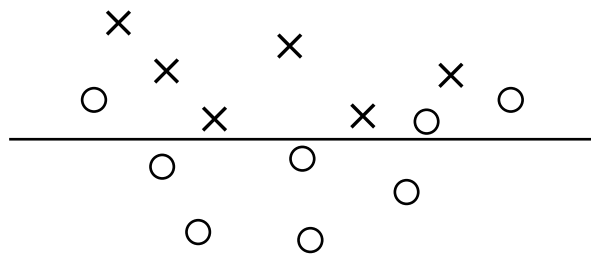
- *Si se dispone de conocimiento de las estadísticas del problema: de modo **paramétrico**, según lo que a ellas convenga.*
- *En caso contrario, ha de procederse de forma **no paramétrica**: recurriendo a una función con buenas características de aproximación (preferiblemente, un aproximador universal).*

Q: ¿Como elegiría el número de parámetros de F ?

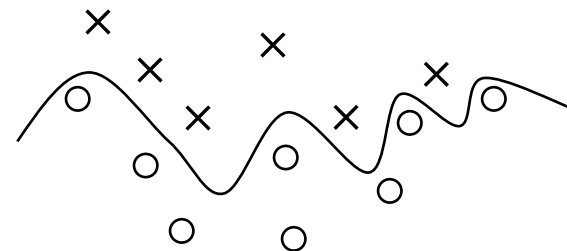
Evitando los fenómenos de

- **subajuste**: *insuficiencia del número de parámetros para encontrar una buena solución;*
- **sobreajuste**: *número de parámetros excesivamente alto, que lleva a una solución que incluye demasiado detalle particular del conjunto de entrenamiento.*

Nótese que el número de parámetros puede venir fijado por F (p. ej., si es lineal, número de variables + 1)



Subajuste



Sobreajuste



Q: Obviamente, un modelo polinómico (en las variables) tiene más poder expresivo que uno lineal, porque es un aproximador universal. ¿Encuentra usted algún inconveniente para su utilización?

La potencia expresiva de un aproximador polinómico crece con su grado. La inclusión de monomios de primer, segundo, tercer, ... grado supone la incorporación de un parámetro por cada uno de ellos, y el número de monomios crece rápidamente con su grado: N de grado 1, $\binom{N+1}{2}$ de grado 2, etc.

*Se produce así lo que se conoce como una **explosión dimensional**: demasiados parámetros en el modelo como para permitir una razonable determinación de los mismos en función de los datos.*

T: Busque y discuta una lista de variables utilizadas para la concesión de un crédito.



T: Otros problemas que se plantean en términos de decisión máquina son:

- *la detección de un avería a partir de sus síntomas*
- *el diagnóstico médico a partir de los síntomas de un paciente*
- *la detección de fraude en operaciones con tarjeta de crédito*

Discuta los aspectos generales de dichos problemas del mismo modo que se ha hecho en el caso de concesión de créditos.



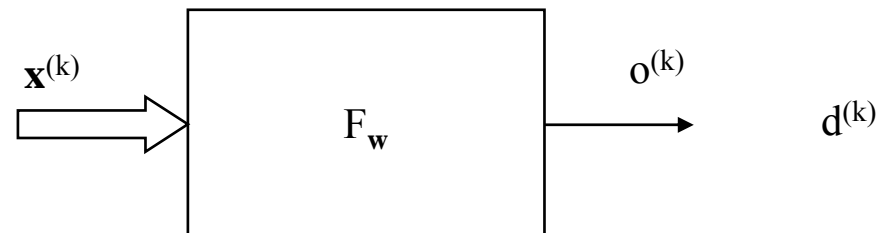
La decisión “blanda”

- En muchos de estos problemas conviene además “**calificar**” la decisión: tener acceso a un resultado continuo que indique no la decisión final (“dura”), sino la probabilidad de mora, o magnitud análoga.
- Tal cosa se puede conseguir mediante un entrenamiento adecuado y la utilización de una función continua F' (como $\mathbf{w}_e^T \mathbf{x}_e$), que después se transforma en F mediante un paso de comparación con un umbral.



Los métodos máquina para decisión

Según se ha visto, consisten en elegir una adecuada F_w y adaptar sus parámetros para optimizar un objetivo sobre las observaciones **etiquetadas** disponibles (entrenamiento supervisado), cuidando de mantener la **generalización**.



$d^{(k)}$ es la etiqueta de la muestra $\mathbf{x}^{(k)}$, o salida **deseada**: si la máquina ofrece salida $o^{(k)}$ y, por ejemplo, se persigue minimizar un **coste** $C(d, o)$, se modifican (en principio secuencialmente) los parámetros \mathbf{w} para reducirlo para cada muestra; utilizando en general, como se ha dicho, un algoritmo de búsqueda.