



# Tema 1. Introducción

Percepción (PER)

Curso 2020/2021

Departamento de Sistemas Informáticos y Computación

- 1 Percepción ▷ 3
- 2 Teoría de la decisión estadística ▷ 7
- 3 Teoría de la decisión estadística en interacción ▷ 13
- 4 Aprendizaje en sistemas interactivos ▷ 17
- 5 Evaluación en sistemas interactivos ▷ 21





- 1 Percepción ▷ 3
  - 2 Teoría de la decisión estadística ▷ 7
  - 3 Teoría de la decisión estadística en interacción ▷ 13
  - 4 Aprendizaje en sistemas interactivos ▷ 17
  - 5 Evaluación en sistemas interactivos ▷ 21





### Percepción

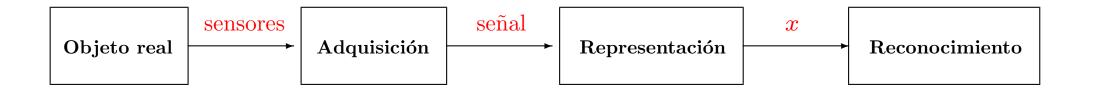
- Percepción como proceso cognitivo para captar información
- La percepción va estrechamente ligada al *Reconocimiento*
- Esta asignatura: Percepción automática y Reconocimiento automático
- Percepción automática como procesamiento informático de la información procedente de sensores
- La información es representada para su tratamiento informático posterior





### Adquisición, Representación y Reconocimiento

Percepción engloba adquisición, representación y reconocimiento



- Sensor: cámara, micrófono, escáner, . . .
- Señal: fichero de imagen, audio, vídeo, texto, . . .
- Representación (x): vector de características, cadena de símbolos, grafo, . . .



#### Reconocimiento

- Reconocimiento puede verse como
  - Clasificación: c(x)=c con  $c\in\{1,\ldots,c,\ldots,C\}$  Ejemplo: determinar categoría gramatical de una palabra x= "clasificar"  $\to c(x)=$  verbo
  - *Interpretación*: varios símbolos relacionados (secuencia, árbol, . . . ) Ejemplo: reconocimiento de texto manuscrito

may be consumed reconocimiento may be consumed

- En esta asignatura veremos reconocimiento como clasificación
- Clasificación desde la teoría de la decisión estadística





- 1 Percepción ▷ 3
- 2 Teoría de la decisión estadística ▷ 7
  - 3 Teoría de la decisión estadística en interacción ▷ 13
  - 4 Aprendizaje en sistemas interactivos ▷ 17
  - 5 Evaluación en sistemas interactivos ▷ 21





#### Teoría de la decisión estadística

La función de clasificador de mínimo error es ascendente, es decir, a mayor probabilidad, debe tener mayor valor.

- lacktriangle La representación del objeto f x a clasificar es una *variable aleatoria*<sup>1</sup>
- Clasificador de Bayes (clasificador de mínimo error):

$$\begin{split} \hat{c}(\mathbf{x}) &= \operatorname*{argmax}_{c=1...C} P(c \mid \mathbf{x}) = \operatorname*{argmax}_{c=1...C} P(\mathbf{x} \mid c) \\ P(\mathbf{x}, \mathbf{c}) &= P(\mathbf{x}) * P(\mathbf{c} \mid \mathbf{x}) \end{split}$$

$$\overset{\text{Regla de Bayes:}}{P(y \mid x)} = \frac{P(y) P(x \mid y)}{P(x)} = \frac{P(y) P(x \mid y)}{\sum_{y'} P(y') P(x \mid y')} \end{split}$$

donde

P(c): **probabilidad a priori** de la clase c

 $p(\mathbf{x} \mid c)$ : distribución de **probabilidad condicionada** a la clase c

- ullet  $p(\mathbf{x} \mid c)$  es una distribución de probabilidad sobre valores de  $\mathbf{x}$  en la clase c
  - Ej: Modelo de Markov oculto (SIN) u otra distribución (Gaussiana, . . . )
  - Se estima a partir de datos de la clase c (entrenamiento)



 $<sup>^1</sup>x$  variable escalar,  $\mathbf x$  variable vectorial

### Ejercicio: clasificador de Bayes

Se dispone de 1000 correos electrónicos etiquetados como Spam (100) y NoSpam (900). Asimismo, la palabra Bingo aparece en 40 correos de los etiquetados como Spam (S) y en 18 de los etiquetados como NoSpam (N).

Dado el siguiente correo:

"Get \$1000 Free - Try the New Slot Machines at Bingo Palace"

¿Cómo lo etiquetaríamos para minimizar la probabilidad de error?





### Ejercicio: clasificador de Bayes

Sea x=1 si la palabra *Bingo* está en el correo. El clasificador de Bayes sería:

$$\hat{c}(x) = \operatorname*{argmax} P(c \mid x = 1) = \operatorname*{argmax} P(c) \, p(x = 1 \mid c)$$
 
$$c \in \{\mathsf{S}, \mathsf{N}\}$$

Estimamos empíricamente las distribuciones de probabilidad involucradas:

$$P(c = S) = \frac{100}{1000}$$

$$P(c = N) = \frac{900}{1000}$$

$$p(x = 1 \mid c = S) = \frac{40}{100}$$

$$p(x = 1 \mid c = N) = \frac{18}{900}$$

Calculamos para cada clase:

$$P(c = S) p(x = 1 \mid c = S) = \frac{100}{1000} \cdot \frac{40}{100} = 0.04$$

$$P(c = N) p(x = 1 \mid c = N) = \frac{900}{1000} \cdot \frac{18}{900} = 0.018$$

Por lo tanto,  $\hat{c}(x) = S$ 





### **Algunas consideraciones**

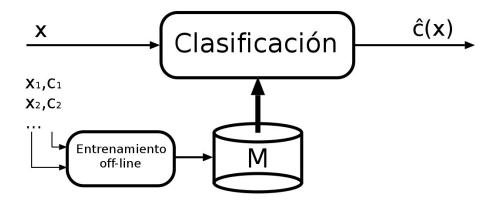
- La variable aleatoria x tendría que dar cuenta de la aparición o ausencia no sólo de la palabra Bingo, sino de todas las palabras del vocabulario
- De esta forma, x podría ser un vector donde cada componente del mismo estuviera asociado a una palabra distinta (Tema 2)
- Este tipo de representación conlleva una alta dimensionalidad que puede ser interesante reducir (Temas 3 y 6)
- Estos objetos pueden compararse por distancias (Tema 4)
- Existe probabilidades condicionales  $p(\mathbf{x} \mid c)$  bien conocidas más allá de la aparición o ausencia de un objeto (palabra) en particular (Tema 5)
- Los objetos de cada clase pueden no ser linealmente separables y podría convenir que lo fueran (Tema 6)
- Podemos usar más de un clasificador y combinarlos (Tema 7)





#### Clasificación con entrenamiento off-line

- Muestras de entrenamiento etiquetadas  $\{(\mathbf{x}_1, c_1), \dots, (\mathbf{x}_n, c_n)\}$
- $\blacksquare$  Estimación de modelo M (P(c) y  $p(\mathbf{x} \mid c)$ ) a partir de entrenamiento
- ullet Clasificación de x de acuerdo a modelo M



■ En esta asignatura solo estudiaremos el entrenamiento off-line





- 1 Percepción ▷ 3
- 2 Teoría de la decisión estadística > 7
- 3 Teoría de la decisión estadística en interacción ▷ 13
  - 4 Aprendizaje en sistemas interactivos ▷ 17
  - 5 Evaluación en sistemas interactivos ▷ 21





#### Teoría de la decisión estadística en interacción

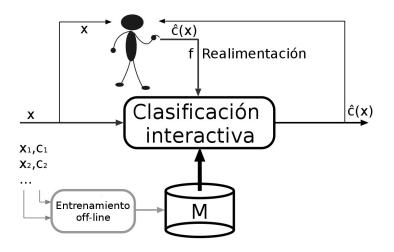
#### Motivación:

- Sistema de reconocimiento (clasificación) puede cometer errores
  - Clasificación de textos, traducción automática, reconocimiento del habla
- Es necesaria la supervisión (interacción) de un usuario (humano)
- Existe supervisión humana en
  - Desarrollo: anotación de los datos empleados en estimar los modelos
  - Producción: corrección de posibles errores
  - Realimentación: considerar las correcciones para mejorar el sistema
- Es necesario adaptar la teoría de la decisión en presencia de interacción



#### Clasificación con realimentación sin reentrenamiento

ullet El usuario realiza una corrección f sobre la clasificación del sistema  $\hat{c}(\mathbf{x})$ 



$$\hat{c}(\mathbf{x}) = \underset{c=1...C}{\operatorname{argmax}} \ P(c \mid \mathbf{x}, f)$$

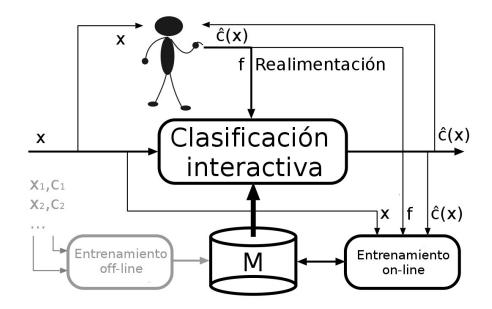
- ullet El modelo M no se modifica, pero el sistema emplea la realimentación f
- Una posible realimentación es solicitar una clasificación alternativa de x





#### Clasificación con realimentación con reentrenamiento

- La realimentación proporciona nuevas muestras de entrenamiento etiquetadas
- ullet Es posible reentrenar el modelo M para incorporar las nuevas muestras



ullet ¿Cómo contribuyen las nuevas muestras al modelo existente M en un sistema interactivo?





- 1 Percepción ▷ 3
- 2 Teoría de la decisión estadística > 7
- 3 Teoría de la decisión estadística en interacción ▷ 13
- 4 Aprendizaje en sistemas interactivos ▷ 17
  - 5 Evaluación en sistemas interactivos ▷ 21





# Aprendizaje en sistemas interactivos

Algunas estrategias de entrenamiento en sistemas interactivos son:

- On-line learning
- Active learning





# **On-line learning**

- Muestras de entrenamiento iniciales  $T = \{(\mathbf{x}_1, c_1), \cdots, (\mathbf{x}_n, c_n)\}$
- La realimentación del usuario genera un nuevo conjunto de entrenamiento  $T' = \{(\mathbf{x}_{n+1}, c_{n+1}), \cdots, (\mathbf{x}_m, c_m)\}$
- El nuevo modelo M' se obtiene de la combinación de T y T', empleando T' para actualizar sus parámetros ("probabilidades")
- lacktriangle Se debe ponderar la contribución entre T y T' a M' mediante interpolación





### **Active learning**

- Se dispone de:
  - Conjunto reducido de muestras etiquetadas  $T = \{(\mathbf{x}_1, c_1), \cdots, (\mathbf{x}_n, c_n)\}$
  - Conjunto amplio de muestras no etiquetadas  $U = \{(\mathbf{x}_{n+1},?), \cdots, (\mathbf{x}_m,?)\}$
- El etiquetado de muestras no etiquetadas es costoso
- El sistema elige  $T' \subset U$  del menor tamaño posible para etiquetar por el usuario
- ullet El error del modelo entrenado con  $T \cup T'$  debe ser menor que con T
- Algunos criterios de selección de muestras a etiquetar:
  - Uncertainty sampling: muestras cuyo etiquetado es más incierto
  - Expected model change: muestras que mayor cambio causan en el modelo





- 1 Percepción ▷ 3
- 2 Teoría de la decisión estadística > 7
- 3 Teoría de la decisión estadística en interacción ▷ 13
- 4 Aprendizaje en sistemas interactivos ▷ 17
- 5 Evaluación en sistemas interactivos ▷ 21





#### Evaluación en sistemas interactivos

- La evaluación convencional basada en tasa de error no es apropiada
- La evaluación se basa en *esfuerzo* o *productividad* del usuario:
  - Esfuerzo: errores corregidos (interacciones requeridas)
  - Productividad: en términos de tiempo requerido para realizar una tarea
- La evaluación automática de un sistema interactivo se basa en esfuerzo
  - Modelo de usuario en base a un conjunto de interacciones (operaciones)
  - Es una aproximación a la función objetivo a maximizar (productividad)
  - Es posible evaluar y comparar sistemas interactivos inmediatamente
- La evaluación manual de un sistema interactivo se basa en productividad
  - Corresponde con la percepción del usuario de la utilidad del sistema
  - Es costosa: reclutar usuarios, organización, control de experimentos, etc.
  - Número de sistemas a evaluar limitado, resultados no inmediatos



