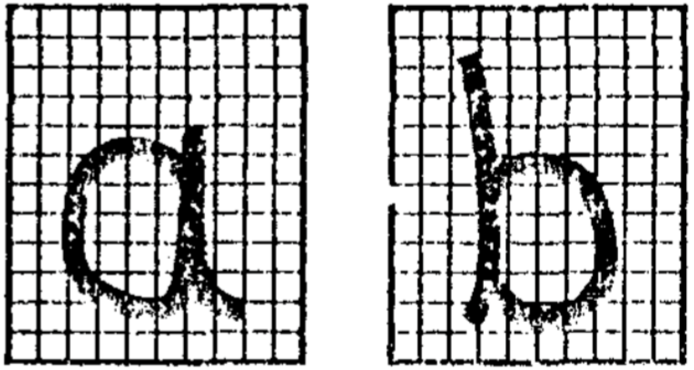
Bayes básico

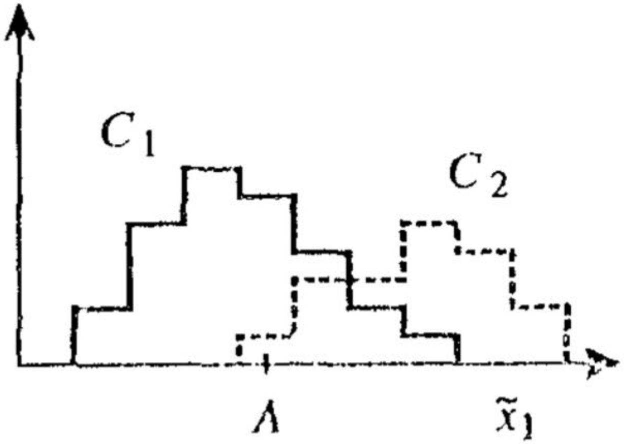
Introducción

Diplomado en minería de datos

Reconocimiento de patrones



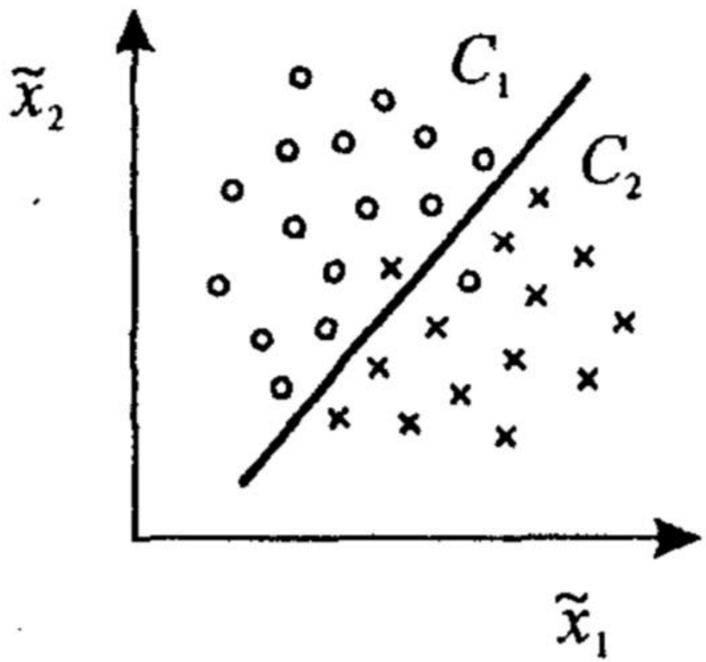
Pre-procesamiento y Extracción de carácterísticas (~)



Altura de la letra

Reconocimiento de patrones

- Agregamos otra característica

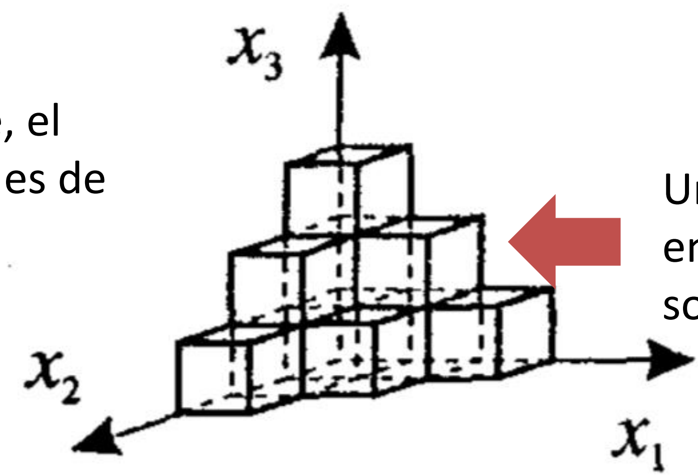


$$y_k = y_k(\vec{x};\vec{w})$$

Maldición de la dimensionalidad

- A partir de cierto punto, agregar nuevas variables lleva a una reducción de desempeño

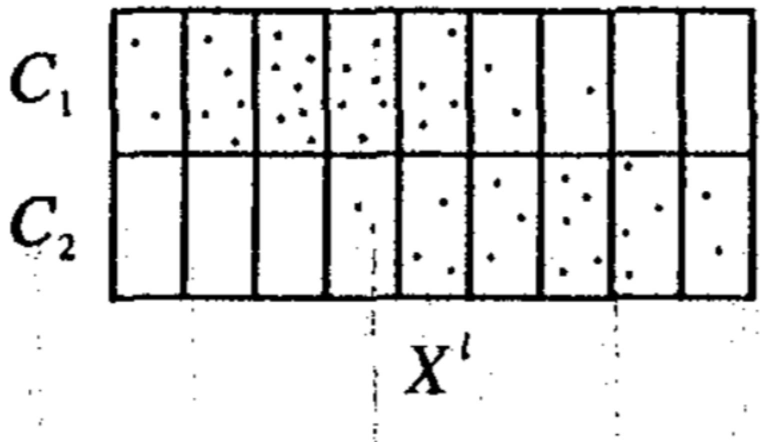
Si subdividimos en M intervalos cada variable, el número total de celdas es de M^d



Uno o varios datos de entrenamiento
son asignados a una celta

Como cada celda debe tener por lo menos un dato de entrenamiento, la cantidad de datos de entrenamiento crece exponencialmente

Clasificador



$$P(C_k) = \frac{\#\ \text{puntos renglón } C_k}{\#\ \text{total de puntos}}$$

$$P(C_k, X^l) = \frac{\#\text{puntos en celda } C_k, X^l}{\#\ \text{total de puntos}}$$

$$P(X^l \mid C_k) = \frac{\#\text{puntos en celda } C_k, X^l}{\#\ \text{puntos renglón } C_k}$$

$$P(C_k, X^l) = P(X^l \mid C_k) P(C_k)$$

$$P(C_k, X^l) = P(C_k \mid X^l) P(X^l)$$

$$P(C_k \mid X^l) P(X^l) = P(X^l \mid C_k) P(C_k)$$

Bayes básico

Probabilidad a-posteriori

Verosimilitud

Probabilidad a-priori

$$P(C_k \mid X^l) = \frac{P(X^l \mid C_k)P(C_k)}{P(X^l)}$$

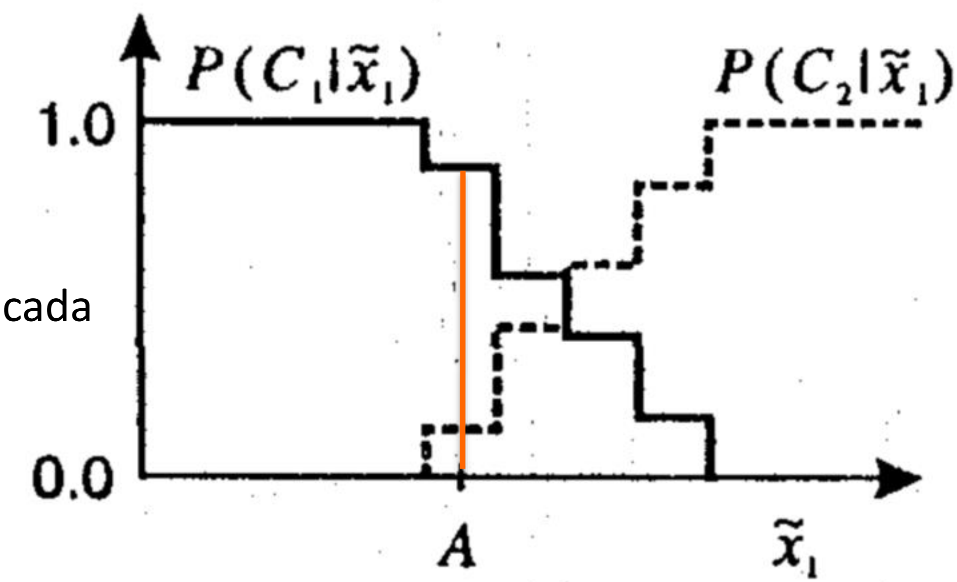
Factor de normalización

$$P(X^l) = \sum_{k=1}^{c} P(X^l \mid C_k)P(C_k)$$

Bayes básico

- Para clasificar una imagen nueva cuya altura es A, minimizando la probabilidad de hacerlo erróneamente, seleccionamos la $P(C_k \mid X^l)$ más grande (regla de Bayes)

A debe ser clasificada como C_1



Bayes básico

- Dos etapas en el proceso de clasificación
  - inferencia: los datos son usados para calcular los valores de las probabilidades a-posteriori
  - toma de decisión: Las probabilidades calculadas en el paso anterior para asignar un datos a alguna de las posibles clases

Bayes ingenuo

- Es un clasificador probabilístico fundamentado en el teorema de Bayes y algunas hipótesis simplificadoras adicionales.
  - hipótesis simplificadoras: independencia de las variables predictoras
  - Es por ello que se le llama Bayes ingenuo

Bayes ingenuo

Escalofríos	Escurrimiento nasal	Dolor de cabeza	Fiebre	Gripa
S	N	Ligero	S	N
S	Y	No	N	S
S	N	Fuerte	S	S
N	Y	Ligero	S	S
N	N	No	N	N
N	Y	Fuerte	S	S
N	Y	Fuerte	N	N
S	Y	Ligero	S	S

Bayes ingenuo

Escalofríos	Escurrimiento nasal	Dolor de cabeza	Fiebre	Gripa
S	N	Ligero	S	N
N	N	No	N	N
N	Y	Fuerte	N	N
S	Y	No	N	S
S	N	Fuerte	S	S
N	Y	Ligero	S	S
N	Y	Fuerte	S	S
S	Y	Ligero	S	S

Datos ordenados por clase de la variable a predecir

Bayes ingenuo

Escalofríos	Escurrimiento nasal	Dolor de cabeza	Fiebre	Gripa
S	N	Ligero	S	N
N	N	No	N	N
N	Y	Fuerte	N	N
S	Y	No	N	S
S	N	Fuerte	S	S
N	Y	Ligero	S	S
N	Y	Fuerte	S	S
S	Y	Ligero	S	S

P(Gripa=N)=3/8=0.375

Bayes ingenuo

Escalofríos	Escurrimiento nasal	Dolor de cabeza	Fiebre	Gripa
S	N	Ligero	S	N
N	N	No	N	N
N	Y	Fuerte	N	N
S	Y	No	N	S
S	N	Fuerte	S	S
N	Y	Ligero	S	S
N	Y	Fuerte	S	S
S	Y	Ligero	S	S

P(Gripa=S)=5/8=0.625

Bayes ingenuo

Escalofríos	Escurrimiento nasal	Dolor de cabeza	Fiebre	Gripa
S	N	Ligero	S	N
N	N	No	N	N
N	Y	Fuerte	N	N
S	Y	No	N	S
S	N	Fuerte	S	S
N	Y	Ligero	S	S
N	Y	Fuerte	S	S
S	Y	Ligero	S	S

P(Gripa=S)+P(Gripa=N)=1

Bayes ingenuo

Escalofríos	Escurrimiento nasal	Dolor de cabeza	Fiebre	Gripa
S	N	Ligero	S	N
N	N	No	N	N
N	Y	Fuerte	N	N
S	Y	No	N	S
S	N	Fuerte	S	S
N	Y	Ligero	S	S
N	Y	Fuerte	S	S
S	Y	Ligero	S	S

P(Escurrimiento Nasal=N, Gripa=N)<=2/8=0.25

Bayes ingenuo

Escalofríos	Escurrimiento nasal	Dolor de cabeza	Fiebre	Gripa
S	N	Ligero	S	N
N	N	No	N	N
N	Y	Fuerte	N	N
S	Y	No	N	S
S	N	Fuerte	S	S
N	Y	Ligero	S	S
N	Y	Fuerte	S	S
S	Y	Ligero	S	S

P(Escurrimiento Nasal=N| Gripa=N)=2/3=0.667

Bayes ingenuo

- Si A y B son 2 eventos independientes:

  $P(A|B)=P(A)$

  $P(A,B)=P(A)P(B)$

Bayes ingenuo

- Quiero conocer si un paciente con los siguientes síntomas tiene o no gripa:

Escalofríos	Escurrimiento nasal	Dolor de cabeza	Fiebre
S	N	Ligero	S

- Debo encontrar el máximo de
  - P(Gripa=S| Escalofríos=S, Escurrimiento nasal=N, Dolor de cabeza=Ligero, Fiebre=S)
  - P(Gripa=N| Escalofríos=S, Escurrimiento nasal=N, Dolor de cabeza=Ligero, Fiebre=S)

Bayes ingenuo

- Probabilidad a-posteriori
  - P(Gripa=S| Escalofríos=S, Escurrimiento nasal=N, Dolor de cabeza=Ligero, Fiebre=S)
- Por el teorema de Bayes tenemos que lo anterior es igual a
  - [P(Escalofríos=S, Escurrimiento nasal=N, Dolor de cabeza=Ligero, Fiebre=S|Gripa=S) x P(Gripa=S)] / P(Escalofríos=S, Escurrimiento nasal=N, Dolor de cabeza=Ligero, Fiebre=S)
- Dado que el factor de normalización es constante y queremos obtener el máximo, lo podemos eliminar (escalarlo no afecta la decisión) y queda
  - P(Escalofríos=S, Escurrimiento nasal=N, Dolor de cabeza=Ligero, Fiebre=S|Gripa=S) x P(Gripa=S)

Bayes ingenuo

- Eso es igual a
  - P(Escalofríos=S, Escurrimiento nasal=N, Dolor de cabeza=Ligero, Fiebre=S, Gripa=S)
- Aplicando la regla de la probabilidad condicional varias veces
  - P(Escalofríos=S| Escurrimiento nasal=N, Dolor de cabeza=Ligero, Fiebre=S, Gripa=S) x P(Escurrimiento nasal=N, Dolor de cabeza=Ligero, Fiebre=S, Gripa=S)
  - P(Escalofríos=S| Escurrimiento nasal=N, Dolor de cabeza=Ligero, Fiebre=S, Gripa=S) x P(Escurrimiento nasal=N| Dolor de cabeza=Ligero, Fiebre=S, Gripa=S) x P(Dolor de cabeza=Ligero, Fiebre=S, Gripa=S)

Bayes ingenuo

- Aplicando la regla de la probabilidad condicional varias veces
  - P(Escalofríos=S| Escurrimiento nasal=N, Dolor de cabeza=Ligero, Fiebre=S, Gripa=S) x P(Escurrimiento nasal=N| Dolor de cabeza=Ligero, Fiebre=S, Gripa=S) x P(Dolor de cabeza=Ligero| Fiebre=S, Gripa=S) x P(Fiebre=S| Gripa=S) x P (Gripa=S)

Bayes ingenuo

- Dado que asumimos que las variables predictoras son independientes entre sí
  - P(Escalofríos=S| Gripa=S) x P(Escurrimiento nasal=N| Gripa=S) x P(Dolor de cabeza=Ligero| Gripa=S) x P(Fiebre=S| Gripa=S) x P (Gripa=S) = 0.006
- Para el Gripa=N tendríamos
  - P(Escalofríos=S| Gripa=N)x P(Escurrimiento nasal=N| Gripa=N)x P(Dolor de cabeza=Ligero| Gripa=N) x P(Fiebre=S| Gripa=N) x P (Gripa=N) = 0.0185
- Es más probable que esos síntomas **no** sea una gripa

Bayes ingenuo

- En general, el algoritmo para determinar como categorizar una serie de variables predictoras sería

$$clasifica(f1,\ldots,fn) = \underset{c}{\arg\max}\, p(C_k = c)\prod_{i=1}^{n} p(F_i = f_i \mid C_k = c)$$

Bayes básico

- Hemos visto ejemplos de variables predictoras discretas
- En varias aplicaciones estás podrían ser continuas
- Las probabilidades de variables discretas se reemplazarían por funciones de densidad de probabilidad

$$P\left(x \in [a,b]\right) = \int_a^b p(x)dx$$

Bayes básico

Probabilidad a-posteriori

Verosimilitud

Probabilidad a-priori

$$P(C_k \mid X^l) = \frac{p(X^l \mid C_k) P(C_k)}{p(X^l)}$$

Factor de normalización

$$p(X^l) = \sum_{k=1}^{c} p(X^l \mid C_k) P(C_k)$$

Bayes básico

$$p\left(x \mid C_k\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)}$$

μ es el promedio de los valores de x asociados con la clase C_k

σ es la desviación estandar de los valores de x asociados con la clase C_k