M2.2.2 Modelos Supervisados y No Supervisados

Programa Big Data y Business Intelligence

Enrique Onieva

enrique.onieva@deusto.es

https://twitter.com/EnriqueOnieva

https://www.linkedin.com/in/enriqueonieva/

Deusto
Facultad de Ingeniería
Ingeniaritza Fakultatea

Naive Bayes

- Motivación
- Probabilidad: ideas intuitivas
- La regla de bayes
- El clasificador Naive Bayes
  - Construcción del modelo
  - Consulta al modelo
  - Extensiones
- Limitaciones y soluciones

Motivación

- Supongamos que vamos al médico y le decimos:
  - "¿Cuál es su diagnóstico?"
    - Posiblemente te responderá "resfriado", o "gastroenteritis"
    - (Aquella enfermedad que sea más común)
  - "Tengo fiebre, ¿Cuál es su diagnóstico?"
    - Responderá aquella la enfermedad más común de aquellas que se manifiestan con fiebre
  - "Tengo fiebre y me duele la cabeza, ¿Cuál es su diagnóstico?"
    - Nuevamente responderá la enfermedad más común de aquellas que se manifiestan con fiebre y dolor de cabeza

Motivación

- En problemas reales (como el caso anterior), sufrimos de incertidumbre
    - Tenemos un conocimiento parcial del estado del problema
        - Sólo conocemos algunos "síntomas"
    - Las observaciones vienen con ruido o errores
    - Puede ocurrir un evento no contemplado por nuestro modelo
    - Es un proceso inherentemente estocástico

Presentación

- Naive Bayes es un método probabilístico
  - Representa conocimiento con incertidumbre.
  - Se puede manipular para razonamiento y toma de decisiones.
  - Se pueden tratar muchas variables.
  - Las probabilidades se pueden estimar a partir de datos.
  - Los modelos tienen una interpretación clara y bien definida.
- SI un paciente tiene fiebre y dolor de cabeza, entonces tiene gripe (certeza 0.7)

Presentación

- La validez de una regla depende del contexto.
  - Si conozco el nivel de estudios de una persona, tengo información sobre su nivel de ingresos.
    - Esta información puede ser equivocada y ponerse de manifiesto si conozco el puesto de trabajo concreto que esta persona desarrolla
  - Si al salir de casa vemos el césped mojado podemos sospechar que ha llovido.
    - Si descubrimos que nos hemos dejado la manguera abierta, dejamos de sospechar que ha llovido.

Presentación

- Las reglas con incertidumbre deberían de poder usarse en ambas direcciones
  - Si hay fuego debe de haber humo
  - Si vemos humo sospechamos la existencia de fuego
- Correlación entre las informaciones.
  - Si una misma información se repite muchas veces no debe de aumentar nuestra certidumbre

Presentación

- Otra forma de expresar el conocimiento: grado de creencia
  - Creemos (por experiencia) que un paciente con congestión nasal, tiene gripe con probabilidad del 70%
  - En teoría de la probabilidad, se expresa como
    - P(Gripe = true | Congestión = true) = 0,7
  - Por tanto, la probabilidad puede cambiar a medida que se conocen nuevas evidencias
- La teoría de la probabilidad servirá como medio de representación del conocimiento incierto

Presentación

- Naive Bayes es un método de clasificación tanto para problemas binarios como multi-clase
- Es Naive (ingenuo) porque asume que todas las entradas son independientes entre sí
  - Que unos atributos no dependen de otros (ni un poco)
- Esto no suele ser así, cuando tratamos con datos reales
- No obstante, funciona "sorprendentemente" bien

Probabilidad: Ideas intuitivas

- Dada una proposición “a”
- Su probabilidad incondicional P(a) cuantifica su grado de creencia
- En ausencia de cualquier otra evidencia
  - Ejemplo: $P(caries) = 0{,}1$
    $P(caries, \neg dolor) = 0{,}05$
    - Notación: $P(caries, \neg dolor) = P(caries \wedge \neg dolor)$
- Aproximación frecuentista: número de casos favorables entre el número de casos totales

Deusto
Facultad de Ingeniería
Ingeniaritza Fakultatea

Probabilidad: Ideas intuitivas

- Distribución de probabilidad
  - Indica las probabilidades de que la variable pueda tomar cada uno de sus valores
  - Ejemplo: Tiempo ={lluvia, sol , nubes, nieve}, su distribución de probabilidad podría ser:
    - P(Tiempo = sol) = 0,7
    - P(Tiempo = lluvia) = 0,2
    - P(Tiempo = nubes) = 0,08
    - P(Tiempo = nieve) = 0,02

Probabilidad: Ideas intuitivas

- Probabilidad Condicional
  - Probabilidad condicional asociada a “a” dado “b”:
    - Grado de creencia sobre a, dado que todo lo que sabemos es que b ocurre → P(a|b)
  - Ejemplo: P(caries|dolor) = 0,8
    - Sabiendo que un paciente tiene dolor de muelas (y sólo sabemos eso), nuestra creencia es que el paciente tendrá caries con probabilidad 0,8
  - Los grados de creencia se actualizan a medida que se van conociendo nuevas evidencias en el mundo incierto
    - P(a|b) = 0,8 no es lo mismo que decir que “siempre que b sea verdad, entonces P(a) = 0,8”
    - Ya que P(a|b) refleja que b es la única evidencia conocida

Probabilidad: Ideas intuitivas

- Independencia probabilística
  - Intuitivamente, dos variables son independientes si conocer el valor que toma una de ellas no nos actualiza nuestra creencia sobre el valor que tome la otra.
    - El asumir que dos variables son independientes está basado normalmente en el conocimiento previo del dominio que se modela
  - Dos variables aleatorias X e Y son independientes si
  - $P(X|Y) = P(X)$
  - $P(Y|X) = P(Y)$
  - $P(X,Y) = P(X) \cdot P(Y)$
- $P(TengoCaries|HaceSol) = P(TengoCaries)$

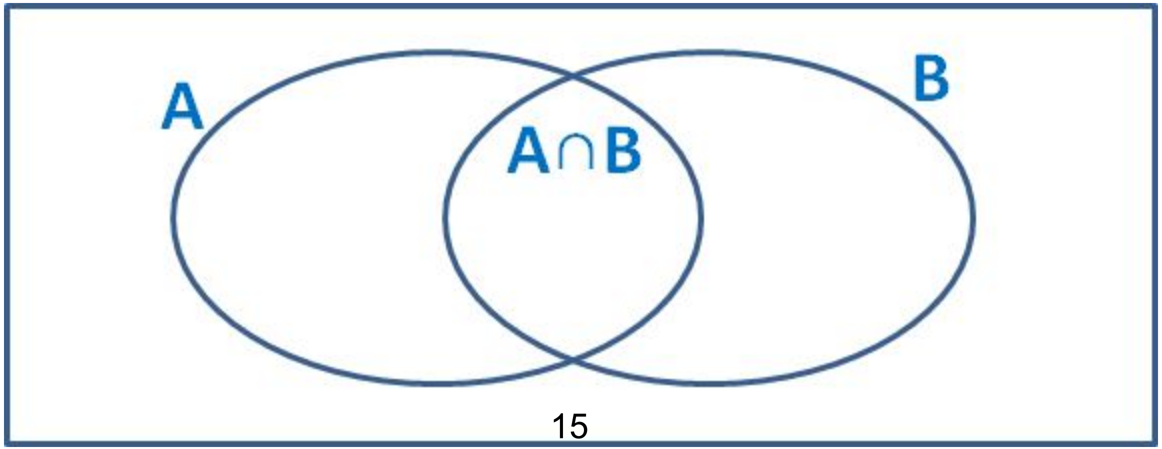
Deusto
Facultad de Ingeniería
Ingeniaritza Fakultatea

La Regla de Bayes

- A partir de
  - $P(a \cap b) = P(a|b)P(b) = P(b|a)P(a)$
- Podemos deducir la siguiente fórmula, conocida como la regla de Bayes
  - $P(\mathrm{b}|a) = (P(a|b)P(b))/(P(a))$
- Regla de Bayes para variables aleatorias:
  - $P(Y|X) = (P(X \cap Y))/(P(X)) = (P(X|Y)P(Y))/(P(X))$
- Generalización, en presencia de un conjunto e de observaciones
  - $P(Y|X,e) = (P(X|Y,e)P(Y|e))/(P(X|e))$

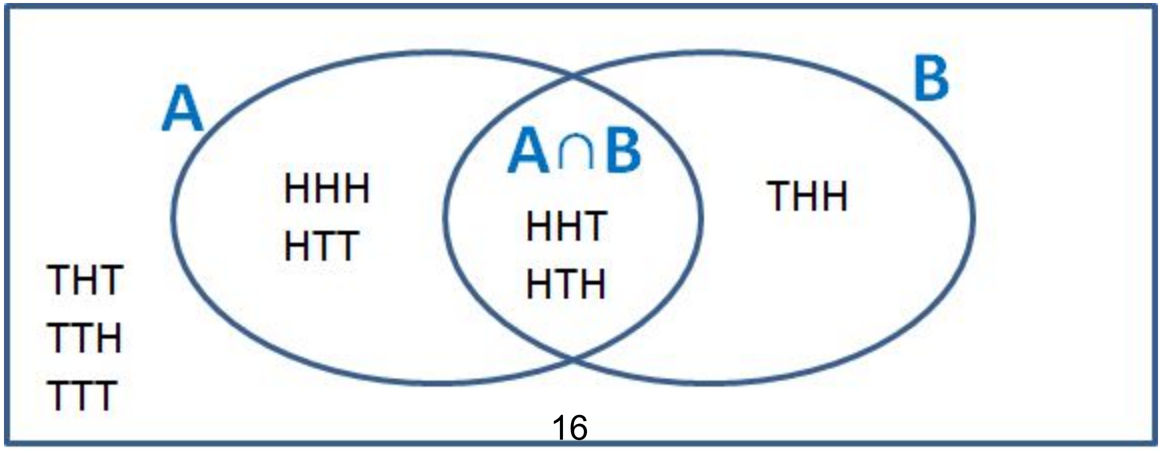
La Regla de Bayes

- Lanzamos una moneda 3 veces. Los posibles casos Cara (Heads, H) y Cruz (Tails, T) son
  - S={HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
- Definimos los eventos:
  - A = “La primera moneda es Cara” = {HHH, HHT, HTH, HTT}
  - B = “Dos de las monedas son Cara”={HHT, HTH, THH}



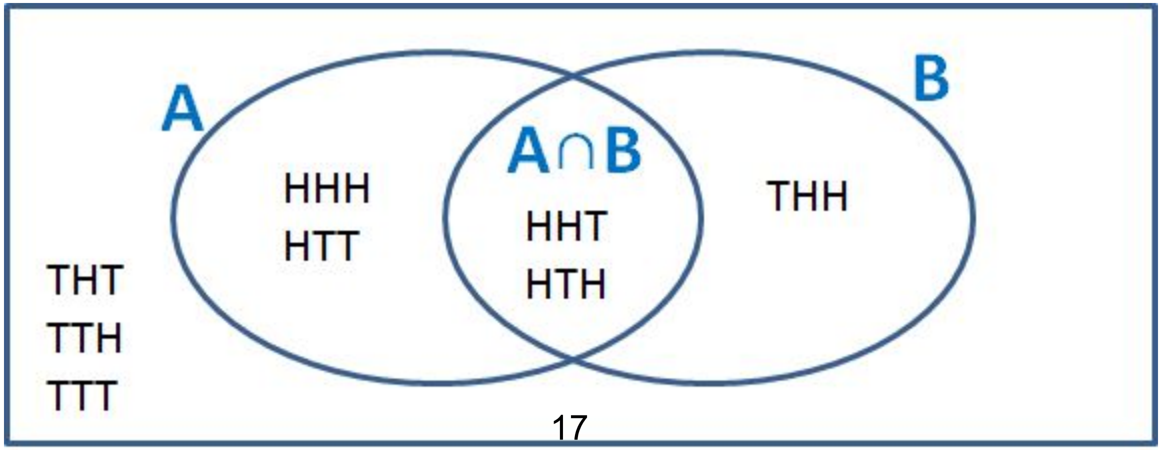
La Regla de Bayes

- Tenemos
    - A = “La primera moneda es Cara” = {HHH, HHT, HTH, HTT} P(A) = 4/8
    - B = “Dos de las monedas son Cara”={HHT, HTH, THH} P(B) = 3/8
- Probabilidad Condicional
    - Tiramos una moneda 3 veces. Si sabemos que hay 2 caras, cuál es la probabilidad de que la primera sea cara?



La Regla de Bayes

- Probabilidad Condicional
  - Tiramos una moneda 3 veces. Si sabemos que hay 2 caras, cuál es la probabilidad de que la primera sea cara?
    - Otra forma: Sabiendo que B es cierto, cuál es la probabilidad de A?
    - B es cierto: hemos obtenido un resultado entre HHT, HTH y THH
    - De esos, aquellos en los que A es cierto son HHT y HTH (A∩B)



Ejemplo “Clásico”

- Un fabricante de caramelos fabrica bolsas de cinco tipos
    - h1: 100% de naranja
    - h2: 75% de naranja y 25% de limón
    - h3: 50% de naranja y 50% de limón
    - h4: 25% de naranja y 75% de limón
    - h5: 100% de limón
  - Cada tipo de bolsa h1, h2, h3, h4 y h5 las hace el fabricante con probabilidad 0,1, 0,2, 0,4, 0,2 y 0,1.

Ejemplo "Clásico"

- Tomamos una bolsa y vamos abriendo algunos caramelos y anotando su sabor ¿Podemos predecir el sabor del siguiente caramelo que saquemos de la bolsa?

Ejemplo “Clásico”

- Supongamos que tomamos 2 caramelos de una bolsa elegida al azar, ambos de limón
- ¿Qué tipo de bolsa hemos tomado?

Hipótesis	P(h)	P(d|h)	P(d|h)·P(h)	α·P(d|h)·P(h)
H1	0,1	0	0	0
H2	0,2	$0{,}25^2$	0,0125	0,04
H3	0,4	$0{,}5^2$	0,1	0,31
H4	0,2	$0{,}75^2$	0,1125	0,35
H5	0,1	1^2	0,1	0,3

Deusto
Facultad de Ingeniería
Ingeniaritza Fakultatea

20

Ejemplo “Clásico”

- Supongamos que tomamos 3 caramelos de una bolsa elegida al azar, ambos de limón
- ¿Qué tipo de bolsa hemos tomado?

Hipótesis	P(h)	P(d|h)	P(d|h)·P(h)	α·P(d|h)·P(h)
H1	0,1	0	0	0
H2	0,2	$0,25^3$	0,003125	0,01
H3	0,4	$0,5^3$	0,05	0,21
H4	0,2	$0,75^3$	0,084375	0,36
H5	0,1	1^3	0,1	0,42

El Clasificador Naive Bayes

- Supongamos un conjunto de atributos $A_1, \ldots A_n$
- Tenemos un conjunto de entrenamiento D con una serie de ejemplos de valores concretos para los atributos, junto con su clasificación
- Queremos aprender un clasificador tal que clasifique nuevas instancias $\langle a_1, \ldots\ a_n \rangle$

Deusto
Facultad de Ingeniería
Ingeniaritza Fakultatea

El Clasificador Naive Bayes

- Podemos simplificar el aprendizaje suponiendo que los atributos son (mutuamente) condicionalmente independientes dado el valor de clasificación (de ahí lo de "naive")
- En ese caso, tomamos como valor de clasificación:
- Algo como:
  - "Aquella hipótesis para la que la probabilidad de los atributos, conociendo la hipótesis, es mayor"

$$v_{NB} = argmax_{v_j \in V} P(v_j) \prod_i P(a_i | v_j)$$

Deusto
Facultad de Ingeniería
Ingeniaritza Fakultatea

El Clasificador Naive Bayes

- Para el proceso de aprendizaje, solo tenemos que estimar las probabilidades:
  - $P(v_j) \rightarrow$ probabilidades a priori
    - Probabilidad de que una muestra sea de una determinada clase, independientemente del valor de sus atributos
  - $P(a_i|v_j) \rightarrow$ probabilidades condicionadas
    - Probabilidad de que una muestra tenga un determinado valor de un atributo, conociendo el valor de la clase
- Simplemente contamos frecuencias
- Esa tabla con las probabilidades, calculadas para los datos de entrenamiento, serán el modelo

Construcción del Modelo

- Probabilidades a priori - $P(v_j)$
  - P(SI) = 9/14 y P(NO) = 5/14
- Probabilidades a posteriori - $P(ai|vj)$
  - P(Cielo=soleado|SI) = 2/9
  - P(Cielo=soleado|NO) = 3/5
  - P(Cielo=nublado|SI) = ...
  - P(Temp=alta|SI) = ...
  - P(Temp=alta|NO) = ...
  - P(Temp=suave|SI) = 5/9
  - P(Temp=suave|NO) = 1/5

EJ	Cielo	Temp.	Humedad	Viento	Jugar Tenis
D_1	Soleado	Alta	Alta	Débil	NO
D_2	Soleado	Alta	Alta	Fuerte	NO
D_3	Nublado	Alta	Alta	Débil	SI
D_4	Lluvia	Suave	Alta	Débil	SI
D_5	Lluvia	Baja	Normal	Débil	SI
D_6	Lluvia	Baja	Normal	Fuerte	NO
D_7	Nublado	Baja	Normal	Fuerte	SI
D_8	Soleado	Suave	Alta	Débil	SI
D_9	Soleado	Baja	Normal	Débil	NO
D_{10}	Lluvia	Suave	Normal	Débil	SI
D_{11}	Soleado	Suave	Normal	Fuerte	SI
D_{12}	Nublado	Suave	Alta	Fuerte	SI
D_{13}	Nublado	Alta	Normal	Débil	SI
D_{14}	Lluvia	Suave	Alta	Fuerte	NO

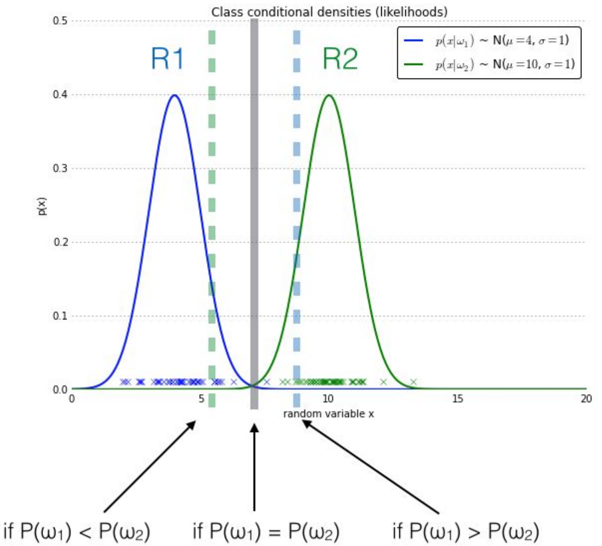
Consultando al Modelo

- Una vez recibimos datos de entrada
  - Si el día es soleado, con temperatura suave, humedad alta y viento fuerte, ¿es un buen día para jugar al tenis?
    - Tomamos de las tablas, aquellas probabilidades relacionadas
      - P(SI)·P(soleado|SI)·P(suave|SI)·P(alta|SI)P(fuerte|SI)=0,011
      - P(NO)·P(soleado|NO)·P(suave|NO)·P(alta|NO)P(fuerte|NO)=0,015
    - Y el clasificador devolverá la clase con mayor valor
      - En este caso la respuesta es NO (no es un día bueno)
    - Si queremos una respuesta con probabilidad, basta con normalizar los resultados
      - SI → 0.011/(0.011+0.015) NO → 0.015/(0.011+0.015)

$P(SI)$	9/14
$P(NO)$	5/14
$P(soleado\|SI)$	2/9
$P(soleado\|NO)$	3/5
$P(suave\|SI)$	5/9
$P(suave\|NO)$	1/5
$P(alta\|SI)$	4/9
$P(alta\|NO)$	3/5
$P(fuerte\|SI)$	3/9
$P(fuerte\|NO)$	3/5

Extensiones

- Para trabajar con datos numéricos
- La probabilidad condicional se calcula como la distribución de probabilidad
  - Media y desviación típica de los valores del atributo para una determinada clase



Aplicación a Text Mining

- Demuestra gran potencial en text mining:
  - Clasificar documentos por tema (economía y tecnología)
  - Bag of words
    - Disponemos de la lista de palabras que aparecen en cada documento
    - Contamos cuántas veces aparece cada palabra en cada documento

	market	stock	price	application	mobile	google
document 1('economics')	1	2	3	0	0	0
document 2('economics')	0	1	2	0	0	1
document 3('technology')	0	0	0	2	3	1
document 4('technology')	1	0	1	2	3	0

Limitaciones y soluciones

- Documento con la palabra “mobile” →
    - P(Eco)·P(market|Eco)·P(stock|Eco)·...·P(mobile|Eco)·P(google|Eco)
    - Pero P(mobile|Eco)=0
  - ¿Un documento con “mobile”, nuca será de economía?
    - ¿Incluso aunque tenga las palabras “market” “stock” “price” repetidas 1000 veces y “application” y “google” no aparezcan?
  - ¿Cómo se podría “solucionar”?

	market	stock	price	application	mobile	google
document 1('economics')	1	2	3	0	0	0
document 2('economics')	0	1	2	0	0	1
document 3('technology')	0	0	0	2	3	1
document 4('technology')	1	0	1	2	3	0

Limitaciones y soluciones

- ¿Ante documento con ninguna de esas palabras?
  - P(Eco)=P(Tec)=0
- ¿Y clasificando entre “economía” y “geología”?
  - Hay muchos más documentos de economía que de geología
    - P(Eco)>>P(Geo)
    - Pero ninguno coindice en palabras
      - P(palabra1|Eco) = P(palabra1|Geo)=P(palabra2|Eco)=P(palabra2|Geo)=...=0

	market	stock	price	application	mobile	google
document 1(’economics’)	1	2	3	0	0	0
document 2(’economics’)	0	1	2	0	0	1
document 3(’technology’)	0	0	0	2	3	1
document 4(’technology’)	1	0	1	2	3	0

30

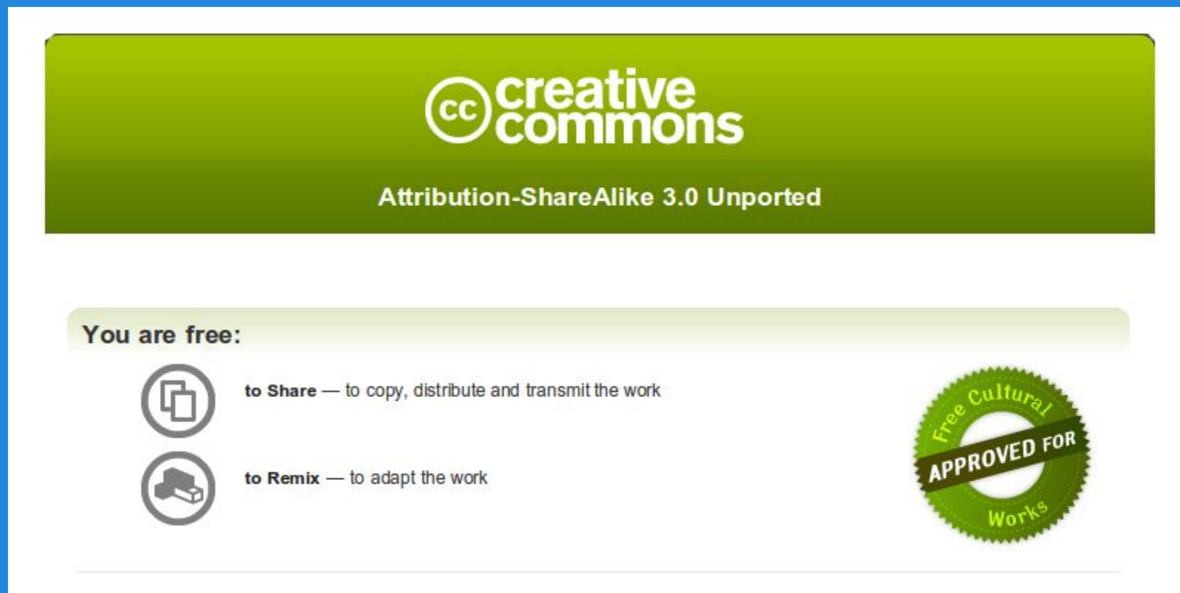
Limitaciones y soluciones

- La corrección de Laplace:
  - Busca evitar esos problemas, haciendo que la probabilidad nunca sea exactamente 0
    - Suma 1 al dividendo y el número de palabras (M) al divisor

$$p(x_i = 1|y = c_k) = \frac{\#\text{ of documents of class } c_k \text{ where word } x_i \text{ appears} + 1}{\#\text{ of documents of class } c_k + M}$$

- Uso de logaritmos:
  - Evitar que los números sean tan infinitamente pequeños que al multiplicarlos el resultado tienda a cero
    - Debido a redondeos del ordenador
    - Calculamos las probabilidades como suma de logaritmos, en lugar de producto de probabilidades

$$\log p(x|y) = \sum_{i=1}^{N} \log p(x_i|y)$$



Enrique Onieva
enrique.onieva@deusto.es
https://twitter.com/EnriqueOnieva
https://www.linkedin.com/in/enriqueonieva/

Deusto
Facultad de Ingeniería
Ingeniaritza Fakultatea