

Método para Classificação:

- Naïve Bayes.

Modelagem Estatística (Bayesiana):

- Contrariamente ao 1R, usa todos os atributos;
- Duas premissas:
 - *Atributos igualmente importantes;*
 - *Atributos estatisticamente independentes* (dada a classe)
 - o valor de um atributo não influencia no valor de outro atributo;
 - Esta premissa quase sempre é violada, mas o método se mostra bastante competitivo na prática. Por quê?
 - Probabilidades estimadas não precisam necessariamente ser corretas, o que importa são as avaliações relativas.
- Tentou o 1R e não ficou satisfeito(a)? Experimente o Naïve Bayes!

Regra de Bayes:

- Probabilidade de um evento H dada a evidência E :

$$\Pr[H | E] = \frac{\Pr[E | H] \Pr[H]}{\Pr[E]}$$

- Probabilidade *a priori* para H : $\Pr[H]$
 - Probabilidade de um evento antes de verificar a evidência
- Probabilidade *a posteriori* para H : $\Pr[H | E]$
 - Probabilidade de um evento após verificar a evidência

Thomas Bayes
(1702-1761)



Naïve Bayes para classificação:

- Qual é a probabilidade da classe dado um exemplo?
 - Evidência E = exemplo (valores dos atributos previsores);
 - Evento H = classe para um exemplo;
- Premissa Naïve: evidência *dividida* em partes (i.e. atributos) independentes.

$$Pr[H | E] = \frac{Pr[E_1 | H] Pr[E_2 | H] \dots Pr[E_n | H] Pr[H]}{Pr[E]}$$

Verossimilhanças para dados "weather":

Outlook			Temperature			Humidity			Windy			Play	
	<i>Yes</i>	<i>No</i>		<i>Yes</i>	<i>No</i>		<i>Yes</i>	<i>No</i>		<i>Yes</i>	<i>No</i>	<i>Yes</i>	<i>No</i>
Sunny	2	3	Hot	2	2	High	3	4	False	6	2	9	5
Overcast	4	0	Mild	4	2	Normal	6	1	True	3	3		
Rainy	3	2	Cool	3	1								
Sunny	2/9	3/5	Hot	2/9	2/5	High	3/9	4/5	False	6/9	2/5	9/14	5/14
Overcast	4/9	0/5	Mild	4/9	2/5	Normal	6/9	1/5	True	3/9	3/5		
Rainy	3/9	2/5	Cool	3/9	1/5								

Outlook	Temp	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Um novo registro:

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

← *Evidência E*

Probabilidade da classe “yes”

$$\begin{aligned} \Pr[\text{yes} \mid E] &= \Pr[\text{Outlook} = \text{Sunny} \mid \text{yes}] \\ &\times \Pr[\text{Temperatur} = \text{Cool} \mid \text{yes}] \\ &\times \Pr[\text{Humidity} = \text{High} \mid \text{yes}] \\ &\times \Pr[\text{Windy} = \text{True} \mid \text{yes}] \\ &\times \frac{\Pr[\text{yes}]}{\Pr[E]} \\ &= \frac{\frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14}}{\Pr[E]} \end{aligned}$$

Continuando...

	Outlook		Temperature		Humidity		Windy		Play				
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No			
Sunny	2	3	Hot	2	2	High	3	4	False	6	2	9	5
Overcast	4	0	Mild	4	2	Normal	6	1	True	3	3		
Rainy	3	2	Cool	3	1								
Sunny	2/9	3/5	Hot	2/9	2/5	High	3/9	4/5	False	6/9	2/5	9/14	5/14
Overcast	4/9	0/5	Mild	4/9	2/5	Normal	6/9	1/5	True	3/9	3/5		
Rainy	3/9	2/5	Cool	3/9	1/5								

- Para um novo dia:

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

Verossimilhança para as duas classes:

$$\text{Para "yes"} = 2/9 \times 3/9 \times 3/9 \times 3/9 \times 9/14 = 0.0053$$

$$\text{Para "no"} = 3/5 \times 1/5 \times 4/5 \times 3/5 \times 5/14 = 0.0206$$

Convertendo para probabilidades por meio de normalização:

$$P(\text{"yes"}) = 0.0053 / (0.0053 + 0.0206) = 0.205$$

$$P(\text{"no"}) = 0.0206 / (0.0053 + 0.0206) = 0.795$$

Problema da frequência *zero*:

- O que acontece se um determinado valor de atributo não aparece na base de treinamento, mas aparece no exemplo de teste?
(e.g. "outlook=overcast" para classe "no")
 - Probabilidade correspondente será zero!
 - *Probabilidade a posteriori* será também zero!
- Possível solução: adicionar 1 ao contador para cada combinação de valor-classe (Estimador de *Laplace*). Como resultado, as probabilidades nunca serão *zero*!

Estimativas das probabilidades modificadas:

- No caso geral, pode-se adicionar uma constante μ diferente de 1;
- Exemplo: atributo *outlook* para a classe *yes*:

$$\frac{2 + \mu / 3}{9 + \mu}$$

Sunny

$$\frac{4 + \mu / 3}{9 + \mu}$$

Overcast

$$\frac{3 + \mu / 3}{9 + \mu}$$

Rainy

Valores ausentes:

- Treinamento: excluir exemplo do conjunto de treinamento;
- Classificação: omitir atributo com valor ausente do cálculo;
- Exemplo:

Outlook	Temp.	Humidity	Windy	Play
?	Cool	High	True	?

Verossimilhança para "yes" = $3/9 \times 3/9 \times 3/9 \times 9/14 = 0.0238$

Verossimilhança para "no" = $1/5 \times 4/5 \times 3/5 \times 5/14 = 0.0343$

Chance ("yes") = $0.0238 / (0.0238 + 0.0343) = 41\%$

Chance ("no") = $0.0343 / (0.0238 + 0.0343) = 59\%$

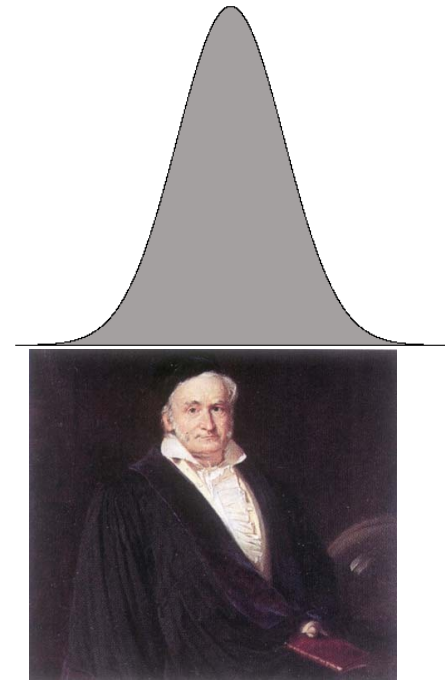
Atributos numéricos:

- Assumir uma distribuição normal para estimar as probabilidades:

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\sigma = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Karl Gauss, 1777-1855

Estatísticas para "weather":

	Outlook		Temperature		Humidity		Windy		Play		
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	
Sunny	2	3	64, 68,	65, 71,	65, 70,	70, 85,	False	6	2	9	5
Overcast	4	0	69, 70,	72, 80,	70, 75,	90, 91,	True	3	3		
Rainy	3	2	72, ...	85, ...	80, ...	95, ...					
Sunny	2/9	3/5	$\mu = 73$	$\mu = 75$	$\mu = 79$	$\mu = 86$	False	6/9	2/5	9/14	5/14
Overcast	4/9	0/5	$\sigma = 6.2$	$\sigma = 7.9$	$\sigma = 10.2$	$\sigma = 9.7$	True	3/9	3/5		
Rainy	3/9	2/5									

- Valor de densidade:

$$f(\text{temperatura} = 66 \mid \text{yes}) = \frac{1}{\sqrt{2\pi} \cdot 6.2} e^{-\frac{(66 - 73)^2}{2 \cdot 6.2^2}} = 0.0340$$

- Qual seria a classe mais provável para o registro

Outlook	Temp.	Humidity	Windy	Play
Sunny	66	83	True	?

Discussão para Naïve Bayes:

- Naïve Bayes funciona bem mesmo quando suas premissas são violadas. Classificação não requer estimativas precisas da probabilidade, desde que a probabilidade máxima seja atribuída à classe correta (Domingos & Pazzani, On the Optimality of the Simple Bayesian Classifier under Zero-One Loss, Machine Learning 29, 103-130, 1997).
- Entretanto, a existência de muitos atributos redundantes pode causar problemas;
- Muitos atributos numéricos não seguem uma distribuição *Gaussiana* (\rightarrow *kernel density estimators*).
- Multiplicação de probabilidades pode causar *underflow* (log)

Extensões para o Naïve Bayes:

- Selecionar melhores atributos (e.g. busca gulosa);
- Redes Bayesianas;

Aplicações usando Naïve Bayes:

- Classificação de textos

- Modelo multinomial

- N = tamanho do texto

$$\Pr[E|H] \approx N! \times \prod_{i=1}^k \frac{P_i^{n_i}}{n_i!}$$

- N_i = número de vezes que o termo ocorreu no documento
- P_i = probabilidade do termo ocorrer na classe H (treinamento)

- Detecção de spam

- Atributos específicos para o domínio;
- Incremental.

Detecção de SPAM - Naïve Bayes:

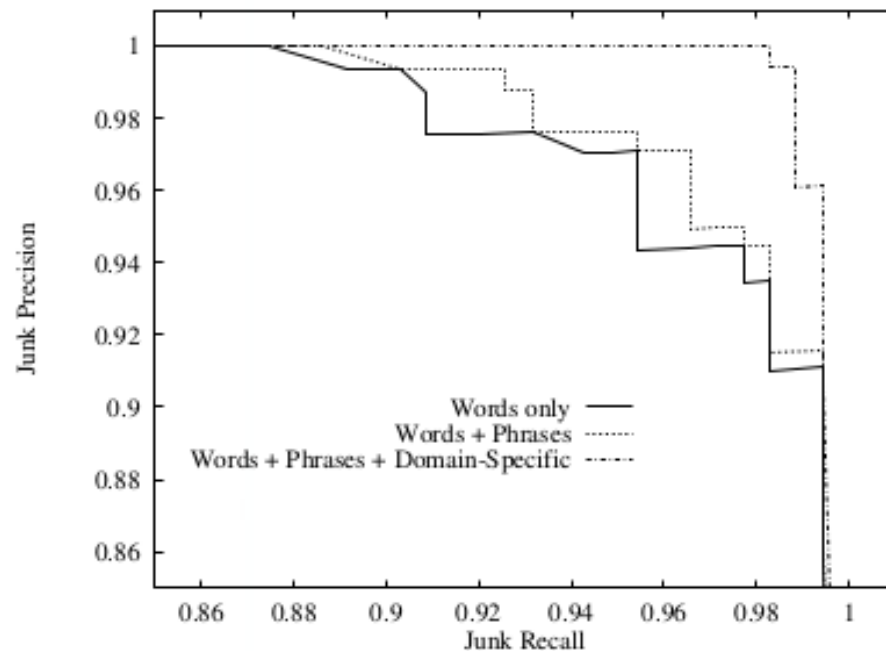


Figure 3: Precision/Recall curves for junk mail using various feature sets.

A Bayesian Approach to Filtering Junk E-Mail

Mehran Sahami*

Susan Dumais[†]

David Heckerman[†]

Eric Horvitz[†]

*Gates Building 1A
Computer Science Department
Stanford University

[†]Microsoft Research
Redmond, WA 98052-6399