

Naive Bayes

Introdução

- Abordagem probabilística (Teorema de Bayes)
- Exemplos
 - Filtros de spam
 - Mineração de emoções
 - Separação de documentos

Base original

História do crédito	Dívida	Garantias	Renda anual	Risco
Ruim	Alta	Nenhuma	< 15.000	Alto
Desconhecida	Alta	Nenhuma	>= 15.000 a <= 35.000	Alto
Desconhecida	Baixa	Nenhuma	>= 15.000 a <= 35.000	Moderado
Desconhecida	Baixa	Nenhuma	> 35.000	Alto
Desconhecida	Baixa	Nenhuma	> 35.000	Baixo
Desconhecida	Baixa	Adequada	> 35.000	Baixo
Ruim	Baixa	Nenhuma	< 15.000	Alto
Ruim	Baixa	Adequada	> 35.000	Moderado
Boa	Baixa	Nenhuma	> 35.000	Baixo
Boa	Alta	Adequada	> 35.000	Baixo
Boa	Alta	Nenhuma	< 15.000	Alto
Boa	Alta	Nenhuma	>= 15.000 a <= 35.000	Moderado
Boa	Alta	Nenhuma	> 35.00000	Baixo
Ruim	Alta	Nenhuma	>= 15.000 a <= 35.000	Alto

Naive Bayes

Risco de crédito	História do crédito			Dívida		Garantias		Renda anual		
	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	< 15000 3	$\begin{matrix} \geq 15000 \\ \leq 35000 \end{matrix}$ 4	≥ 35000 7
	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6
	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3
	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5

Vantagens x desvantagens

- Vantagens

- Rápido
- Simplicidade de interpretação
- Trabalha com altas dimensões
- Boas previsões em bases pequenas

- Desvantagem

- Combinação de características (atributos independentes) – cada par de características são independentes – nem sempre é verdade

Conclusão