

Machine Learning com Python

Prof. Luciano Galdino

NAIVE BAYES

Um dos primeiros algoritmos em Machine Learning.

Classificador probabilístico baseado na aplicação do teorema de Bayes.

$$P(A | B) = \frac{P(B | A) P(A)}{P(B)}$$

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

Premissa: independência entre as variáveis do problema.

Trabalha muito bem com variáveis categóricas.

Algumas Aplicações

Filtros de spam.

Diagnósticos médicos.

Classificação de informações textuais.

Análise de crédito.

Separação de documentos.

Previsão de falhas.

Vantagens:

- Rápido e de fácil entendimento.
- Pouco esforço computacional.
- Bom desempenho com muitos dados.
- Boas previsões com poucos dados

Desvantagens:

- Considera atributos independentes.
- Atribuição de um valor nulo de probabilidade quando uma classe contida no conjunto de teste não se apresenta no conjunto de treino.