

Revisão da prova

Marcos Vinicius - T17 - G1

Dissertativa 1

Perguntas

1. O que é acurácia em um modelo de classificação? Explique o que significa uma acurácia de 100% no conjunto de treinamento.
2. Qual a diferença entre a acurácia no conjunto de treinamento e a acurácia no conjunto de teste?
3. Qual é o problema que pode ocorrer ao obter uma acurácia de 100% no conjunto de treinamento com Naive Bayes? Explique como você chegou a essa conclusão e quais características do modelo podem indicar esse problema.
4. Como suposição de independência das características no Naive Bayes pode influenciar o desempenho do modelo? O que pode acontecer se essa suposição não for verdadeira, especialmente no contexto de análise de spam?

Respostas

1. Acurácia é a porcentagem de instâncias classificadas corretamente pelo modelo. Uma acurácia de 100% no conjunto de treinamento, indica que o modelo aprendeu exatamente como funcionam os dados deste conjunto.
2. A acurácia de 100% no conjunto de teste prevê que o modelo se adaptou perfeitamente os dados daquele conjunto, portanto, para ele será mais difícil interpretar dados novos, como os do conjunto de teste.
3. Como mencionado antes, devido a alta adaptação aos dados já conhecidos, o modelo tem dificuldade em interpretar novos dados, por isso, encontramos aqui um caso de Overfitting. É possível de perceber isso graças a acurácia perfeita nos dados de treinamento e os valores distoantes entre treino e teste
4. O Naive Bayes é um algoritmo de classificação, que trata cada coluna como independente para classificar os dados. É um algoritmo muito bom quando se trata de performance, consegue aprender rápido e até mesmo com uma baixa quantidade de dados. É um algoritmo utilizado muito comumente com análise de spam, por sua habilidade de aprender com poucos dados e conseguir trabalhar bem com PLN (Processamento de Linguagem Natural).

Dissertativa 2

Perguntas Pergunta 1