



As principais causas do Overfitting e Underfitting

☰ Ciclo	Ciclo 04: As garantias de aprendizado
# Aula	32
🕒 Created	@February 24, 2023 8:48 AM
☑ Done	☑
☑ Ready	☑

Objetivo da Aula:

- ☐ Overfitting
- ☐ Underfitting
- ☐ Resumo
- ☐ Próxima aula

Conteúdo:

▼ 1. Overfitting

▼ 1.1 As principais causas do Overfitting

1. **Complexidade do Modelo:** modelos muito complexos podem se ajustar demais aos dados de treinamento e ter mau desempenho em novos dados.
2. **Conjunto de Dados Pequeno:** modelos podem aprender padrões aleatórios nos dados de treinamento quando o conjunto de dados é muito pequeno.

3. **Treinamento Excessivo:** quando o modelo é treinado por muitas épocas, pode se ajustar demais aos dados de treinamento e não generalizar bem.
4. **Vazamento:** informações dos dados de teste vazam para o modelo durante o treinamento, o que pode levar a um ajuste excessivo aos dados de teste.

▼ 1.2 Soluções para o Overfitting

1. **Validação cruzada:** dividir o conjunto de dados em conjuntos de treinamento e validação para avaliar o desempenho do modelo em novos dados.
2. **Regularização:** adicionar uma penalidade à função de perda durante o treinamento para evitar que o modelo se ajuste demais aos dados de treinamento.
3. **Early stopping:** interromper o treinamento do modelo quando o desempenho no conjunto de validação começa a piorar, evitando que o modelo se ajuste demais aos dados de treinamento.
4. **Aumentar o tamanho do conjunto de dados:** aumentar o tamanho do conjunto de dados pode ajudar o modelo a aprender padrões mais relevantes e evitar o ajuste excessivo aos dados de treinamento.

▼ 2. Underfitting

▼ 2.1 As principais causas do Underfitting

1. **Modelo muito simples:** modelos muito simples não conseguem capturar a complexidade dos dados e podem ter um desempenho inferior.
2. **Dados insuficientes:** quando o conjunto de dados é muito pequeno, o modelo pode não ter informação suficiente para aprender padrões relevantes.
3. **Falta de treinamento suficiente:** o modelo pode não ter sido treinado por tempo suficiente para aprender padrões relevantes nos dados.
4. **Modelo mal projetado:** um modelo mal projetado pode ter dificuldade em capturar as nuances dos dados e, portanto, ter um desempenho inferior.

▼ 2.2 Soluções para o Underfitting

1. **Aumentar a complexidade do modelo:** aumentar a complexidade do modelo pode ajudar a capturar a complexidade dos dados e melhorar o desempenho.
2. **Adicionar mais features:** adicionar mais recursos ao conjunto de dados pode fornecer informações adicionais para o modelo aprender.
3. **Aumentar o tempo de treinamento:** treinar o modelo por mais tempo pode ajudar a capturar padrões mais complexos nos dados.
4. **Alterar os parâmetros do modelo:** alterar o design do modelo pode ajudar a capturar nuances adicionais nos dados e melhorar o desempenho.

▼ 3. Resumo

1. O overfitting e o underfitting são os maiores riscos que um algoritmo treinado pode assumir ao ser publicado em Produção.
2. Os Cientistas de Dados precisam estar atentos as causas do overfitting e underfitting e aplicar correções quando necessário.

▼ 4. Próxima aula

Validação Holdout na prática com Python