Treinamento do modelo

Treinar um modelo de aprendizado de máquina com o conjunto de dados de treinamento (train\_data) geralmente envolve os seguintes passos:

1. **Preparação dos dados:** Antes de treinar o modelo, é essencial preparar os dados. Isso pode incluir etapas como **limpeza de dados, normalização, tratamento de valores ausentes e codificação de variáveis categóricas**.
2. **Seleção do algoritmo de aprendizado:** Escolha o algoritmo de aprendizado adequado para o tipo de problema que você está enfrentando (classificação, regressão, agrupamento, etc.) e os requisitos específicos do seu projeto.
3. **Treinamento do modelo:** Use o conjunto de dados de treinamento (train\_data) para ajustar os parâmetros do modelo ao padrão nos dados. Isso envolve alimentar os dados ao modelo e permitir que ele aprenda a relação entre as variáveis de entrada (características) e a variável de destino (rótulo).
4. **Avaliação do modelo:** Após o treinamento, é importante avaliar o desempenho do modelo usando o conjunto de dados de teste (test\_data). Isso geralmente envolve prever os rótulos para os dados de teste e comparar essas previsões com os rótulos reais para calcular métricas de avaliação, como precisão, recall, F1-score, erro quadrático médio, etc.
5. **Ajuste do modelo:** Com base nos resultados da avaliação, você pode ajustar hiperparâmetros do modelo, experimentar diferentes algoritmos ou realizar outras otimizações para melhorar o desempenho do modelo.

O conjunto de dados de teste (test\_data) é usado apenas para avaliar o desempenho do modelo após o treinamento. Não é usado para ajustar os parâmetros do modelo, pois isso poderia resultar em superajuste (overfitting) ao conjunto de dados de teste.

Em resumo, o conjunto de dados de treinamento é usado para ensinar o modelo a reconhecer padrões nos dados, enquanto o conjunto de dados de teste é usado para avaliar o quão bem o modelo generaliza para novos dados não vistos.