

Avaliação de Modelos e Métricas de Qualidade

Avaliação de Modelos

Após a criação (treino) de um modelo usando uma técnica de aprendizagem (machine learning), é necessário avaliar o seu desempenho. A medição do desempenho de um modelo é feita com dados não apresentados durante o treino, para garantir que o modelo é capaz de generalizar para novos dados.

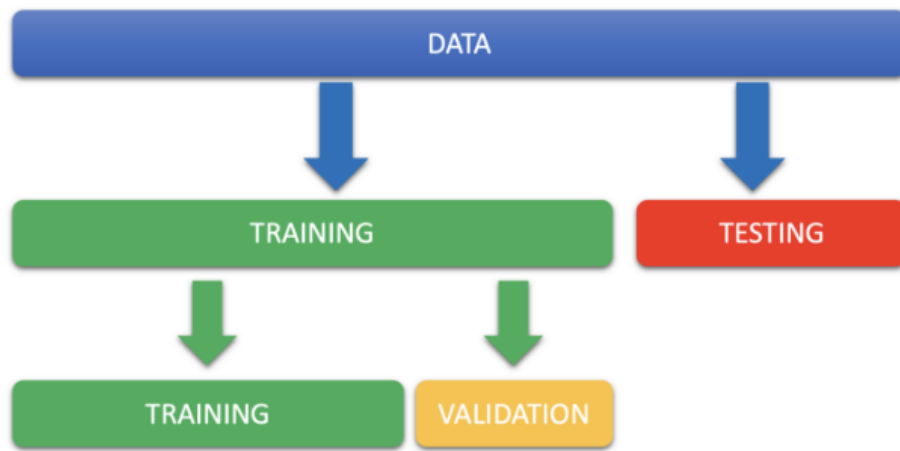


Figure 1: Avaliação de Modelos

Dados de Treino: Conjunto de dados usado para treinar o modelo.

Dados de Teste: Conjunto de dados usados para fornecer uma avaliação imparcial de um modelo final ajustado.

Dados de Validação: Conjunto de dados usados para fornecer uma avaliação imparcial de um ajuste do modelo, no conjunto de dados de treino.

Cross-Validation

O **Cross-Validation** é uma técnica usada para avaliar a capacidade de generalização de um modelo. Ela consiste, utilizando várias iterações, em dividir o conjunto de dados em subconjuntos de treino e teste, treinar o modelo em cada

subconjunto de treino e avaliá-lo em cada subconjunto de teste. O erro final é dado pela média dos valores parciais dos erros.

Leave-One-Out Cross-Validation

Esta técnica de validação cruzada é similar à anterior, mas com uma diferença: em cada iteração, é usado para teste apenas um único exemplo do conjunto de dados.

Qual é o número ideal de k (folds)? - Se o dataset for grande, um valor pequeno para k pode ser suficiente, uma vez que teremos uma quantidade grande de dados para treino. - Se o dataset for pequeno, um valor grande de k (aproximado de N) pode revelar-se mais adequado para maximizar a quantidade de dados de treino. - Quanto maior a quantidade de *folds*, maior a estimativa do erro, mais baixo será o *bias* e menor será o sobreajuste (*overfitting*).

Métricas de Qualidade

As **Métricas de Qualidade** são usadas para avaliar o desempenho de um modelo de aprendizagem automática. No entanto, ela depende também do problema em mãos.

Matriz de Confusão

A **Matriz de Confusão** é uma tabela que permite a visualização do desempenho de um algoritmo de classificação. Ela é composta por quatro campos: - **Verdadeiro Positivo (VP)**: O modelo previu corretamente a classe positiva. - **Falso Positivo (FP)**: O modelo previu incorretamente a classe positiva. - **Verdadeiro Negativo (VN)**: O modelo previu corretamente a classe negativa. - **Falso Negativo (FN)**: O modelo previu incorretamente a classe negativa.

Accuracy

A **Accuracy** é a métrica mais comum para avaliar a qualidade de um modelo de classificação. Ela é calculada pela razão entre o número de previsões corretas e o número total de previsões.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision

A **Precision** é uma medida de exatidão que determina a proporção de itens relevantes entre todos os itens.

$$Precision = \frac{TP}{TP + FP}$$

Recall

O **Recall** é uma medida de completude que determina a proporção de itens relevantes que foram obtidos.

$$Recall = \frac{TP}{TP + FN}$$

Referências

- Cross-Validation