

Dia 2

Módulo 4: Avaliação e Confiabilidade de Modelos

Métricas, Erros e Interpretabilidade Química

Métricas, Erros e Interpretabilidade Química:

Métricas de Classificação (Acurácia, F1, ROC).

Métricas de Regressão (R^2 , RMSE).

O Dilema do Aprendizado: **Overfitting vs. Underfitting**.

Prática: **Matriz de Confusão e Importância de Features**.

O Modelo é Confiável?

Treinar um modelo é fácil, garantir que ele funcione em dados novos é o desafio.

Generalização: Capacidade do modelo de acertar em moléculas que nunca viu antes (Conjunto de Teste).

Cenário Químico:

- Um **Falso Positivo** em toxicidade pode descartar um bom fármaco.
- Um **Falso Negativo** pode levar uma molécula perigosa para testes clínicos.

A Matriz de Confusão

Compara o Preditivo vs. Real.

Verdadeiros Positivos (VP):

Modelo diz que é ativo e é ativo.

Verdadeiros Negativos (VN):

Modelo diz que é inativo e é inativo

Falsos Positivos (FP): "Alarme falso" (Erro Tipo I).

Falsos Negativos (FN):

"Oportunidade perdida" (Erro Tipo II).

		Detectada	
		Sim	Não
Real	Sim	Verdadeiro Positivo (VP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Traduzindo a Matriz em Números

Acurácia: Taxa global de acerto. (Cuidado: enganosa em dados desbalanceados!).

Precisão: Dos que chamei de ativos, quantos realmente são? (Evita Falsos Positivos).

Recall (Sensibilidade): De todos os ativos existentes, quantos eu encontrei? (Evita Falsos Negativos).

F1-Score: Média harmônica entre Precisão e Recall. O melhor equilíbrio.

Avaliando a Robustez

ROC (Receiver Operating Characteristic): Gráfico de Taxa de VP vs. Taxa de FP. Mostra o trade-off do modelo.

AUC (Area Under Curve):

- 0.5: Aleatório (jogar uma moeda).
- 1.0: Perfeito.
- > 0.7: Geralmente aceitável para QSAR.

Interpretação: Probabilidade de o modelo classificar um ativo aleatório acima de um inativo aleatório.

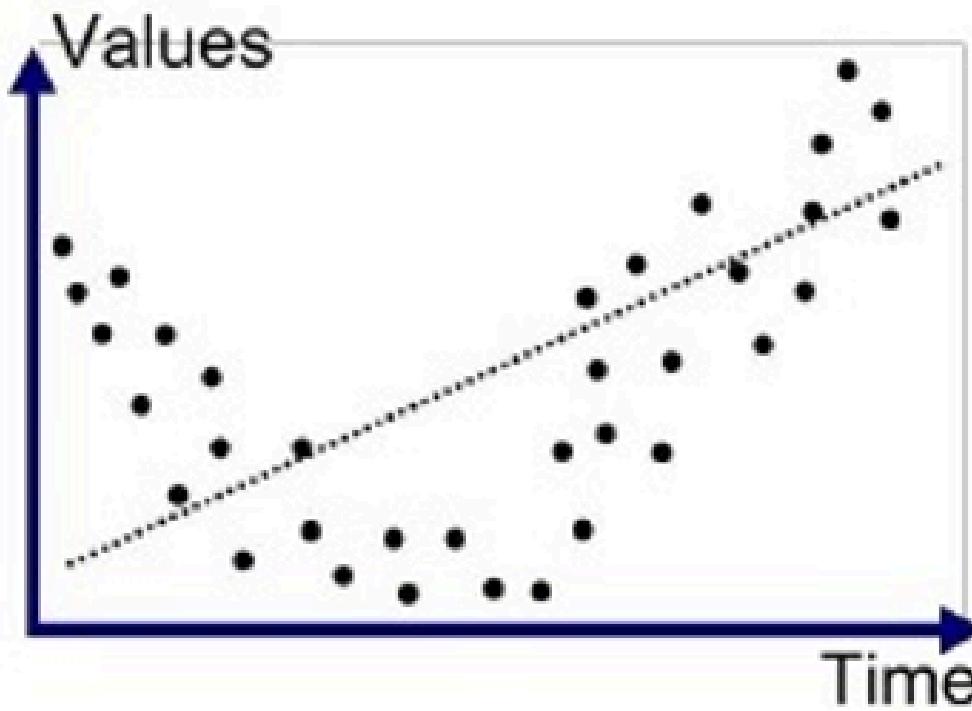
E se o alvo for um número? (Ex: Energia)

Quando prevemos **valores contínuos**, a **Matriz de Confusão não se aplica**.

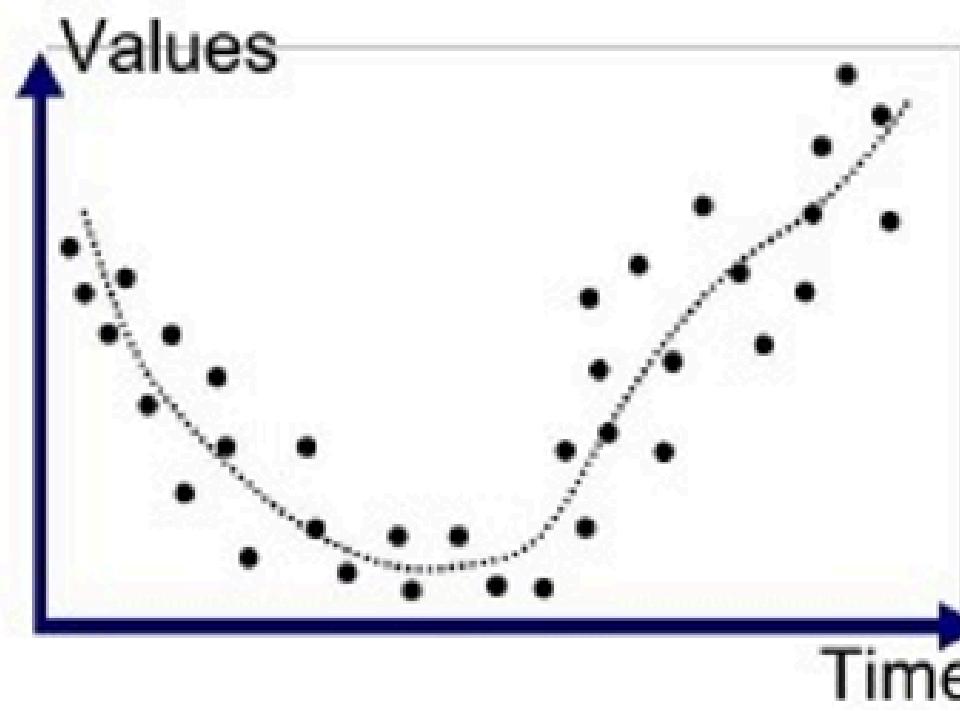
RMSE (Root Mean Squared Error): Erro médio na mesma unidade da propriedade. Penaliza grandes erros.

R² (Coeficiente de Determinação): O quanto da variação dos dados é explicado pelo modelo. (Quanto mais próximo de 1.0, melhor).

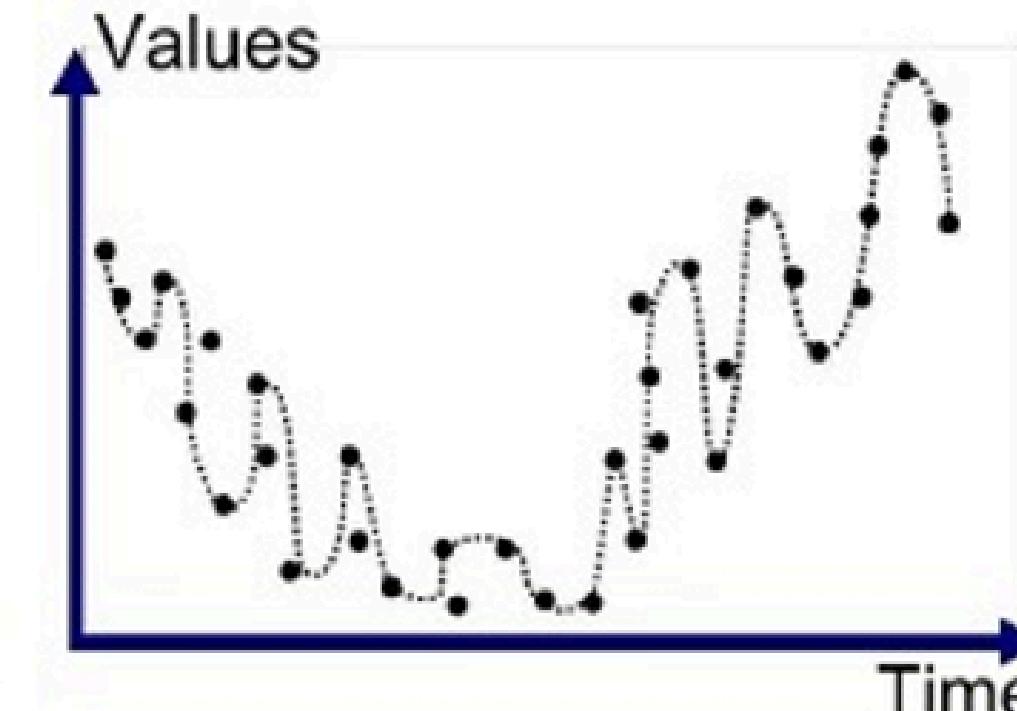
Problemas de Fitting



Underfitted



Good Fit/Robust



Overfitted

Prática: Avaliação Completa

1. Carga: Importar o dataset HIV.
2. Análise: Verificar a proporção de Ativos vs. Inativos.
3. Features (X): Gerar Morgan Fingerprints (converter SMILES em bits).
4. Split Estratificado: Dividir Treino/Teste mantendo a proporção original de ativos (stratify).
5. Treino Ponderado: Treinar Random Forest com ajuste de pesos.
6. Métricas: Analisar Precisão e Recall (ignorando a Acurácia enganosa).
7. Visualização: Gerar Matriz de Confusão e Curva ROC-AUC.