MINERAÇÃO DE DADOS

Prof. Me. Napoleão Póvoa Ribeiro Filho

UNITINS

GOVERNO DO TOCANTINS
TRABALHANDO E CUIDANDO DE TODOS

REGRESSÃO

- Tipo de problema de aprendizado de máquina
- Tem por objetivo **prever valores numéricos contínuos**
- Baseia-se em padrões aprendidos a partir de dados históricos

UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

QUANDO USAR

- Quando o resultado é uma quantidade e não uma classe
- Quando precisamos estimar ou prever tendências

EXEMPLOS

- Previsão de preço de casas
- Previsão de temperatura
- Variação de ações na bolsa
- Estimativa de consumo de energia

UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

PRINCIPAIS ALGORITMOS

- Regressão Linear
- Regressão Polinomial
- Árvore de Regressão
- Random Forest Regressor
- Gradient Boosting
- SVR (Support Vector Regression)

MÉTRICAS DE AVALIAÇÃO

Erro Absoluto Médio - MAE (Mean Absolute Error)

- Calcula a média das diferenças absolutas entre valores reais e previstos
- É fácil de entender, pois está na mesma escala da variável-alvo
- Utilizar quando todos os erros têm impacto igual (sem penalizar extremos)
- Vantagem: menos sensível a outliers
- MAE ≥ 0 (zero é perfeito; quanto menor, melhor)

$$MAE = \frac{1}{n} \sum |y - \hat{y}|$$

UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

MÉTRICAS DE AVALIAÇÃO

Erro Quadrático Médio - MSE (Mean Squared Error)

- Os erros são elevados ao quadrado
- Penaliza fortemente os erros grandes
- Utilizar quando falsos previsores grandes são críticos (ex.: engenharia, saúde)
- MSE $\geq$ 0 (zero é perfeito; sem limite superior)

$$MSE = \frac{1}{n}\sum(y - \hat{y})^2$$

UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

MÉTRICAS DE AVALIAÇÃO

Raiz do Erro Quadrático Médio - RMSE (Root Mean Squared Error)

- Semelhante ao MSE, mas recoloca o erro na mesma escala da variável
- Vantagem: é mais interpretável
- Utilizar para avaliação global da qualidade da previsão
- RMSE ≥ 0 (zero é perfeito; sem limite superior)

$$RMSE = \sqrt{MSE}$$

UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

MÉTRICAS DE AVALIAÇÃO

R^2 (Coeficiente de Determinação)

- Indica quanto da variabilidade dos dados foi explicada pelo modelo
  - 1 → predição perfeita (valor máximo)
  - 0 → modelo não explica nada melhor que média
  - Valores negativos → pior que prever sempre a média
  - Quanto maior, melhor
- Quando usar: para avaliar capacidade explicativa

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

MÉTRICAS DE AVALIAÇÃO

R^2 Ajustado

- Inclui penalização pelo número de features
- Evita overfitting em modelos com muitas variáveis
- Quando usar: modelos múltiplos (várias features)
- Varia sem limite inferior (pode ficar bem negativo)
- Máximo = 1

$$R^2_{adj} = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1}$$

UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

MÉTRICAS DE AVALIAÇÃO

MAPE (Mean Absolute Percentage Error)

- Indica o erro médio percentual
- Excelente para comunicação com público não técnico
- Desvantagem: explode quando valores reais são próximo de 0.
  - MAPE < 10% → excelente
  - 10 – 20% → bom
  - 20 – 50% → aceitável
  - > 50% → ruim

$$MAPE = \frac{100}{n} \sum \left| \frac{y - \hat{y}}{y} \right|$$

UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

MÉTRICAS DE AVALIAÇÃO

Mediana do Erro Absoluto (MdAE)

- Mais robusta a outliers do que o MAE
- Foca em erro típico ao invés da média
- MdAE ≥ 0 (quanto menor, melhor)

MÉTRICAS DE AVALIAÇÃO

MSLE (Mean Squared Logarithmic Error)

- Compara log de valores reais e previstos
- Bom para dados com crescimento exponencial (vendas, epidemias)
- Penaliza menos quando os erros são grandes, mais proporcionais
- MSLE $\geq$ 0 (zero é perfeito; sem limite superior)

RESUMO

Métrica	Escala da saída	Sensível a outliers	Intervalo
MAE	SIM	NÃO	[0, +∞)
MSE	NÃO	SIM	[0, +∞)
RMSE	SIM	SIM	[0, +∞)
R^2	-	Moderado	(-∞, 1]
R^2 Ajustado	-	Moderado	(-∞, 1]
MAPE	%	SIM	[0%, +∞)
MSLE	-	NÃO	[0, +∞)
MdAE	SIM	NÃO	[0, +∞)

UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

QUANDO USAR QUAL

- **MAE** → Quando queremos saber, em média, **o quanto o modelo erra** na prática, sem exagerar a importância dos erros muito grandes.
- **MSE / RMSE** → Quando **erros grandes são mais graves** e precisamos penalizá-los com mais força (ex.: engenharia, saúde, orçamento).
- **R^2** → Quando queremos saber **o quanto o modelo consegue explicar dos dados**. Ajuda a ver o “poder explicativo” do modelo.
- **R^2 Ajustado** → Quando **temos muitas variáveis** e queremos ver se elas realmente ajudam ou se só aumentam o risco de overfitting.

QUANDO USAR QUAL

- **MAPE** → Quando queremos **mostrar o erro em percentual (%)**, de forma simples para pessoas leigas (ex.: clientes, gestores).
- **MSLE** → Quando os valores **crescem muito rápido** (exponencial). Evita punir erros proporcionais e é útil para previsões de crescimento.
- **MdAE** → Quando há outliers (valores fora do normal) e **queremos uma medida de erro mais robusta**, menos influenciada por esses pontos extremos.

PRINCIPAIS DESAFIOS

- Overfitting / Underfitting
- Outiliers
- Milticolinearidade
- Escala dos dados


UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS