

Aula 3 – Regressão Linear e Logística: A Base dos Modelos

Bem-vindo à terceira etapa da sua jornada no mundo da modelagem preditiva! Se você já se perguntou como as empresas preveem vendas futuras, como os bancos avaliam o risco de crédito ou como os sistemas de saúde estimam a probabilidade de uma doença, você está no lugar certo. A resposta, em muitos casos, reside nos fundamentos que exploraremos hoje: a Regressão Linear e a Regressão Logística.

Esses modelos são como as fundações de um grande edifício: talvez não sejam os mais chamativos, mas são absolutamente essenciais para a estabilidade e o sucesso de qualquer construção mais complexa. Compreender seus princípios não só solidifica sua base em Machine Learning, mas também abre portas para interpretar e construir modelos preditivos mais avançados com confiança.

Ao final desta aula, você será capaz de diferenciar e aplicar os princípios da Regressão Linear (simples e múltipla) para prever valores contínuos, interpretar seus coeficientes e avaliar a performance do modelo usando métricas como R^2 , MSE e MAE. Além disso, você será introduzido à Regressão Logística, dominando a função sigmoide, o conceito de log-odds e a matriz de confusão para resolver problemas de classificação. Prepare-se para desvendar a lógica por trás das previsões e classificações que moldam nosso dia a dia digital.

Regressão Linear Simples: A Linha Que Preve

Imagine que você está tentando entender a relação entre o tempo de estudo de um aluno e a nota que ele tira em uma prova. Intuitivamente, esperamos que quanto mais o aluno estuda, maior será sua nota, certo? A Regressão Linear Simples é exatamente isso: uma ferramenta estatística que nos ajuda a modelar a relação entre duas variáveis, onde uma delas (a variável dependente) é explicada pela outra (a variável independente).



Conceito-chave: A Regressão Linear Simples representa a relação entre duas variáveis através de uma linha reta, permitindo visualizar conexões e fazer previsões precisas.

Essa relação é representada por uma linha reta, a famosa "linha de melhor ajuste". Pense nela como uma bússola que aponta a direção geral da relação entre seus dados. Ela nos permite não apenas visualizar essa conexão, mas também fazer previsões. Se soubermos o tempo de estudo de um novo aluno, podemos usar essa linha para estimar sua provável nota.

A beleza da Regressão Linear Simples reside em sua clareza e interpretabilidade. Ela nos dá uma equação que descreve essa linha, permitindo quantificar o impacto da variável independente na dependente. É o ponto de partida para qualquer análise preditiva que envolva dados numéricos contínuos.

Regressão Linear Múltipla: Adicionando Complexidade ao Real

A vida raramente é tão simples que uma única variável explique tudo. Voltando ao exemplo da nota do aluno, além do tempo de estudo, outros fatores como a qualidade do material didático, o número de horas de sono antes da prova ou até mesmo a participação em aulas extras podem influenciar o resultado. É aqui que a Regressão Linear Múltipla entra em cena, permitindo-nos considerar a influência de diversas variáveis independentes sobre uma única variável dependente.

Regressão Simples

Uma variável independente

Linha em 2D

Regressão Múltipla

Múltiplas variáveis independentes

Plano ou hiperplano multidimensional

Em vez de uma linha em um gráfico 2D, pense em um plano ou um hiperplano em um espaço multidimensional. Embora seja difícil visualizar além de três dimensões, a lógica permanece a mesma: encontrar a "superfície" que melhor se ajusta aos dados, minimizando a distância entre os pontos de dados reais e os pontos previstos por essa superfície. Isso nos dá um modelo mais robusto e realista, capaz de capturar a complexidade do mundo real.

A Regressão Linear Múltipla é uma ferramenta poderosa para entender como múltiplos fatores interagem para influenciar um resultado. Ela nos permite isolar o efeito de cada variável, controlando o impacto das outras, o que é crucial para tomadas de decisão informadas em cenários complexos, como prever o preço de um imóvel com base em sua área, número de quartos, localização e ano de construção.

Entendendo os Coeficientes: A Voz dos Números

Quando construímos um modelo de regressão, seja ele simples ou múltiplo, o resultado é uma equação. Essa equação contém números muito importantes chamados **coeficientes**. Eles são a alma do modelo, pois nos dizem não apenas a direção da relação entre as variáveis, mas também a magnitude dessa influência.

O que são coeficientes?

Pense nos coeficientes como os "pesos" que cada variável independente tem sobre a variável dependente. Por exemplo, em um modelo que prevê o preço de uma casa, o coeficiente associado ao número de quartos pode ser 50.000.

Isso significa que, *mantendo todas as outras variáveis constantes* (uma premissa importante conhecida como *ceteris paribus*), cada quarto adicional está associado a um aumento de R\$ 50.000 no preço da casa.

A interpretação correta dos coeficientes é fundamental. Ela transforma números abstratos em *insights acionáveis*. Sem essa compreensão, o modelo seria apenas uma "caixa preta" que cospe previsões. Com ela, podemos entender o "porquê" por trás das previsões e tomar decisões estratégicas, como quais características de um produto impactam mais suas vendas ou quais fatores de risco são mais críticos para uma doença.

Coeficiente de Intercepto

O coeficiente de intercepto representa o valor esperado da variável dependente quando todas as variáveis independentes são zero.



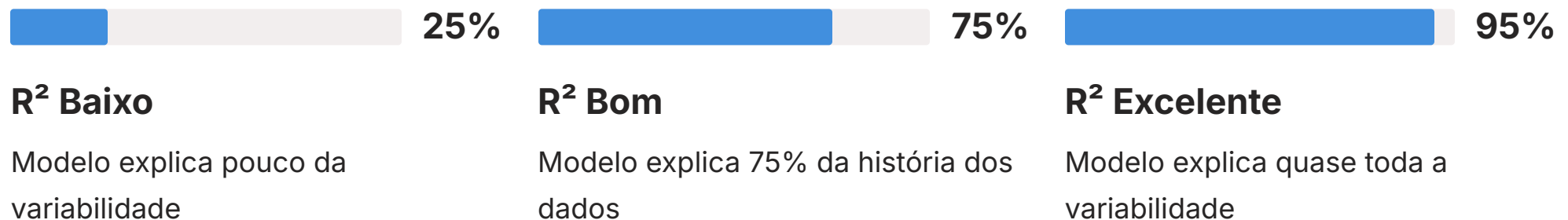
Importante: A interpretação correta dos coeficientes transforma números abstratos em *insights acionáveis*.

Avaliando o Modelo: Será Que Ele Funciona Bem? (Parte 1)

Métrica Principal: R-quadrado (R^2)

Construir um modelo é apenas metade do caminho; a outra metade, igualmente crucial, é saber se ele realmente funciona bem. Como podemos ter certeza de que as previsões do nosso modelo são confiáveis? Para isso, usamos métricas de avaliação. Uma das mais conhecidas para modelos de regressão é o **R-quadrado (R^2)**.

O R^2 nos dá uma ideia de quão bem o nosso modelo explica a variabilidade da variável dependente. Imagine que você está tentando prever a altura de várias pessoas. Sem nenhum modelo, sua melhor "previsão" seria a altura média de todas as pessoas. O R^2 nos diz o quanto nosso modelo é melhor do que essa "previsão média". Ele varia de 0 a 1 (ou 0% a 100%), onde 0 significa que o modelo não explica nada da variabilidade, e 1 significa que ele explica 100% (um cenário ideal e raro na prática).



Um R^2 de 0.75, por exemplo, indica que 75% da variação na variável dependente pode ser explicada pelas variáveis independentes do nosso modelo. É como se o nosso modelo conseguisse contar 75% da história por trás dos dados. Embora um R^2 alto seja geralmente desejável, é importante lembrar que ele não é o único critério e não garante que o modelo seja bom para previsões em novos dados, pois pode ser inflacionado por um número excessivo de variáveis.

Avaliando o Modelo: Será Que Ele Funciona Bem? (Parte 2)

Métricas de Erro: MSE e MAE

Embora o R^2 seja um excelente indicador da capacidade explicativa do modelo, ele não nos diz diretamente o quão "erradas" são nossas previsões em termos absolutos. Para isso, precisamos de métricas que quantifiquem o erro médio do modelo. Duas das mais comuns são o **Erro Quadrático Médio (MSE - Mean Squared Error)** e o **Erro Absoluto Médio (MAE - Mean Absolute Error)**.

MSE - Erro Quadrático Médio

Como funciona: Pega cada erro, eleva ao quadrado e calcula a média

Característica: Penaliza erros grandes mais severamente

Quando usar: Quando erros grandes são particularmente indesejáveis



MAE - Erro Absoluto Médio

Como funciona: Pega o valor absoluto de cada erro e calcula a média

Característica: Mais robusto a outliers

Quando usar: Para uma medida direta do erro médio em unidades da variável

Pense no erro como a diferença entre o valor que seu modelo previu e o valor real observado. O MSE pega cada um desses erros, eleva ao quadrado (para eliminar sinais negativos e penalizar erros maiores mais severamente), e depois calcula a média. Já o MAE simplesmente pega o valor absoluto de cada erro e calcula a média.

  **Objetivo:** Ambas as métricas buscam ser o mais próximo possível de zero, indicando que as previsões do modelo estão muito próximas dos valores reais.

A escolha entre MSE e MAE depende do contexto. O MSE é mais sensível a erros grandes devido à sua penalização quadrática, o que pode ser útil quando erros grandes são particularmente indesejáveis. O MAE, por outro lado, é mais robusto a *outliers* (valores extremos) e fornece uma medida mais direta do erro médio em unidades da variável dependente. Ambas as métricas buscam ser o mais próximo possível de zero, indicando que as previsões do modelo estão muito próximas dos valores reais.

Regressão Logística: Saindo do Contínuo para a Classificação

Até agora, falamos sobre prever valores contínuos, como preços ou notas. Mas e se o nosso objetivo não for prever um número, mas sim uma *categoria*? Por exemplo, queremos saber se um cliente vai comprar um produto (Sim/Não), se um e-mail é spam ou não spam, ou se um paciente tem uma doença ou não. Nesses casos, a Regressão Linear não é a ferramenta adequada, pois ela pode gerar previsões fora do intervalo de 0 e 1, que não fazem sentido para probabilidades ou categorias binárias.

Regressão Linear

- Prevê valores contínuos
- Saída: qualquer número real
- Exemplo: preço, temperatura, nota
- Linha reta

Regressão Logística

- Prevê categorias (classificação)
- Saída: probabilidade entre 0 e 1
- Exemplo: Sim/Não, Spam/Não-spam
- Curva em forma de S

É aqui que a **Regressão Logística** entra em jogo. Apesar do nome "regressão", ela é fundamentalmente um modelo de *classificação*. Em vez de prever um valor diretamente, ela estima a *probabilidade* de um evento ocorrer. Essa probabilidade é então usada para classificar a observação em uma das categorias. Se a probabilidade for maior que um certo limiar (geralmente 0.5), classificamos como "Sim"; caso contrário, como "Não".

Imagine que você é um porteiro que decide se alguém pode entrar ou não. Você não dá uma nota de 0 a 100, mas sim uma decisão binária. A Regressão Logística atua como esse porteiro, mas com a nuance de que ela primeiro calcula a "chance" de entrada (a probabilidade) e só então toma a decisão final. É a base para muitos sistemas de classificação que encontramos diariamente.

A Função Sigmoidal: A Curva Mágica da Probabilidade

O Coração da Regressão Logística

Como a Regressão Logística consegue transformar qualquer valor em uma probabilidade entre 0 e 1? A resposta está na **função sigmoide**, também conhecida como função logística. Esta função é a peça central que permite à Regressão Logística lidar com problemas de classificação.

01

Entrada

Qualquer número real (de $-\infty$ a $+\infty$)

02

Transformação

Aplicação da função sigmoide (curva em S)

03

Saída

Probabilidade entre 0 e 1

Pense na função sigmoide como um "compressor" ou "normalizador" de valores. Ela pega qualquer número real (de menos infinito a mais infinito) e o "espreme" para dentro do intervalo de 0 a 1. O resultado é uma curva em forma de "S" que é suave e contínua. Valores muito negativos se aproximam de 0, valores muito positivos se aproximam de 1, e o valor 0 é mapeado para 0.5.

📌 ✨ **Mágica da Sigmoidal:** A Regressão Logística primeiro calcula uma combinação linear das variáveis de entrada (assim como na regressão linear), e então passa esse resultado pela função sigmoide. O valor de saída da sigmoide é a probabilidade estimada de que a observação pertença à classe positiva.

Essa transformação é crucial porque as probabilidades, por definição, devem estar entre 0 e 1. A Regressão Logística primeiro calcula uma combinação linear das variáveis de entrada (assim como na regressão linear), e então passa esse resultado pela função sigmoide. O valor de saída da sigmoide é a probabilidade estimada de que a observação pertença à classe positiva. É essa "mágica" que permite que um modelo de regressão seja usado para classificação.

Log-Odds: A Linguagem Interna da Regressão Logística

Para entender a Regressão Logística mais a fundo, precisamos mergulhar no conceito de **log-odds**. Enquanto a função sigmoide nos dá a probabilidade final (entre 0 e 1), o que a Regressão Logística realmente modela internamente é o *logaritmo das chances* (log-odds).

1	2	3
Probabilidade	Chances (Odds)	Log-Odds
P = 0.8 (80% de chance)	$0.8 / (1-0.8) = 4$ 4 vezes mais chances de acontecer	$\ln(4) = 1.39$ Pode variar de $-\infty$ a $+\infty$

As "chances" (odds) de um evento são a razão entre a probabilidade de ele acontecer e a probabilidade de ele não acontecer. Por exemplo, se a probabilidade de um cliente comprar é de 0.8 (80%), as chances são de $0.8 / (1 - 0.8) = 0.8 / 0.2 = 4$. Isso significa que o cliente tem 4 vezes mais chances de comprar do que de não comprar. O log-odds é simplesmente o logaritmo natural dessas chances.

A vantagem de trabalhar com log-odds é que eles podem variar de menos infinito a mais infinito, assim como os resultados de uma regressão linear. Isso permite que a Regressão Logística use uma estrutura linear para modelar algo que, após a aplicação da sigmoide, se torna uma probabilidade. Os coeficientes da Regressão Logística, portanto, são interpretados como o impacto de uma unidade de mudança na variável independente sobre o log-odds da variável dependente. É uma forma elegante de conectar a simplicidade da regressão linear com a necessidade de prever probabilidades.

A Matriz de Confusão: Desvendando os Acertos e Erros da Classificação

Depois de treinar um modelo de classificação, como a Regressão Logística, precisamos saber o quão bem ele está se saindo. A **Matriz de Confusão** é uma ferramenta fundamental para avaliar o desempenho de um classificador, especialmente em problemas de classificação binária (duas classes). Ela nos oferece uma visão detalhada de onde o modelo acertou e onde errou.

Imagine que você tem um modelo que prevê se um e-mail é spam ou não. A matriz de confusão organiza as previsões do seu modelo em quatro categorias principais, comparando-as com os valores reais:

Verdadeiro Positivo (VP)

O modelo previu "spam" e o e-mail *era* spam.

✓ **Acerto**

Verdadeiro Negativo (VN)

O modelo previu "não spam" e o e-mail *não era* spam.

✓ **Acerto**

Falso Positivo (FP)

O modelo previu "spam", mas o e-mail *não era* spam.

× **Erro tipo I**

Falso Negativo (FN)

O modelo previu "não spam", mas o e-mail *era* spam.

× **Erro tipo II**

Esta matriz é como um placar detalhado que nos permite entender não apenas o número total de acertos, mas também os tipos específicos de erros que o modelo está cometendo. É crucial para decidir se o modelo é adequado para a tarefa, pois em alguns contextos, um tipo de erro pode ser muito mais custoso do que outro.

Além da Matriz de Confusão: Métricas Derivadas

A matriz de confusão é a base, mas a partir dela podemos calcular métricas mais específicas que nos dão uma visão mais granular do desempenho do nosso modelo de classificação. Essas métricas são essenciais porque, dependendo do problema, podemos priorizar diferentes aspectos do desempenho.

Acurácia (Accuracy)

Fórmula: $(VP + VN) / (VP + VN + FP + FN)$

Proporção de previsões corretas sobre o total de previsões. Intuitiva, mas pode ser enganosa em datasets desbalanceados.

Precisão (Precision)

Fórmula: $VP / (VP + FP)$

Das previsões positivas do modelo, quantas estavam realmente corretas? Importante quando o custo de um Falso Positivo é alto.

Exemplo: diagnosticar alguém com uma doença grave que não tem.

Recall (Sensibilidade)

Fórmula: $VP / (VP + FN)$

Das instâncias realmente positivas, quantas o modelo conseguiu identificar? Importante quando o custo de um Falso Negativo é alto.

Exemplo: não diagnosticar alguém que tem uma doença grave.

F1-Score

Fórmula: Média harmônica da Precisão e do Recall

Útil quando você precisa de um equilíbrio entre Precisão e Recall, especialmente em datasets desbalanceados.

📌 **Escolha Estratégica:** A escolha da métrica mais relevante depende diretamente do problema de negócio. Em um sistema de detecção de fraudes, por exemplo, um alto Recall pode ser mais importante para garantir que poucas fraudes passem despercebidas, mesmo que isso signifique alguns Falsos Positivos.

A escolha da métrica mais relevante depende diretamente do problema de negócio. Em um sistema de detecção de fraudes, por exemplo, um alto Recall pode ser mais importante para garantir que poucas fraudes passem despercebidas, mesmo que isso signifique alguns Falsos Positivos. Compreender essas métricas permite uma avaliação mais estratégica e alinhada aos objetivos do projeto.

Tendência 1: AutoML - Automatizando a Construção de Modelos

O Futuro da Ciência de Dados

No cenário atual da ciência de dados, a velocidade e a eficiência são cruciais. É nesse contexto que o **AutoML (Automação de Machine Learning)** ganha destaque. Imagine ter um assistente inteligente que pode automatizar grande parte do processo de ponta a ponta da aplicação de machine learning, desde o pré-processamento dos dados até a seleção e otimização dos modelos. Isso é o AutoML.

01

Limpeza de Dados

Automatizada

02

Engenharia de Features

Automatizada

03

Seleção de Algoritmos

Automatizada

04

Ajuste de Hiperparâmetros

Automatizado

05

Validação

Automatizada

Tradicionalmente, construir um modelo de ML envolve muitas etapas manuais e iterativas: limpeza de dados, engenharia de features, seleção de algoritmos, ajuste de hiperparâmetros e validação. O AutoML visa simplificar e acelerar essas tarefas, permitindo que cientistas de dados e até mesmo usuários com menos experiência construam modelos de alta performance de forma mais rápida e com menos esforço.

Plataformas Populares

- Google Cloud AutoML
- H2O.ai
- DataRobot
- Auto-Sklearn

Benefícios

- Democratização do ML
- Velocidade na construção
- Foco no problema de negócio
- Escalabilidade

Plataformas e bibliotecas como Google Cloud AutoML, H2O.ai, DataRobot e Auto-Sklearn são exemplos de como essa tendência está se materializando. Elas democratizam o acesso ao machine learning, permitindo que as equipes se concentrem mais na formulação do problema e na interpretação dos resultados, em vez de ficarem presas aos detalhes técnicos da construção do modelo. É uma ferramenta poderosa para escalar a aplicação de ML em diversas indústrias.

Tendência 2: XAI (Explainable AI) - Entendendo o Porquê das Decisões

Transparência e Confiança em IA

À medida que os modelos de Machine Learning se tornam mais complexos e poderosos, especialmente com o avanço de redes neurais e algoritmos de *gradient boosting*, surge uma questão fundamental: como podemos entender e confiar nas decisões que eles tomam? A **Inteligência Artificial Explicável (XAI - Explainable AI)** é a resposta a essa pergunta.

O Problema

Modelos complexos funcionam como "caixas pretas" - fazem previsões sem explicar o porquê.

A Solução: XAI

Técnicas que fornecem insights sobre como e por que um modelo chegou a uma conclusão.

O Resultado

Transparência, confiança e capacidade de validar decisões da IA.

A XAI foca na interpretabilidade dos modelos. Em vez de aceitar uma previsão como uma "caixa preta", ela busca fornecer *insights* sobre como e por que um modelo chegou a uma determinada conclusão. Isso é vital em áreas reguladas, como finanças e saúde, onde a transparência e a capacidade de justificar uma decisão são requisitos legais e éticos. Por exemplo, um banco precisa explicar por que um empréstimo foi negado, e um médico precisa entender por que um sistema de IA sugeriu um tratamento específico.

Técnicas Populares de XAI:

- **SHAP (SHapley Additive exPlanations):** Desvendam a contribuição de cada feature para uma previsão específica
- **LIME (Local Interpretable Model-agnostic Explanations):** Criam explicações locais para previsões individuais

Técnicas como SHAP (SHapley Additive exPlanations) e LIME (Local Interpretable Model-agnostic Explanations) são ferramentas populares de XAI que ajudam a desvendar a contribuição de cada *feature* para uma previsão específica. Elas transformam modelos complexos em algo mais compreensível, construindo confiança e permitindo que os usuários validem e aprimorem o comportamento da IA. A XAI é um pilar para o desenvolvimento de uma IA mais responsável e transparente.

Regressão na Prática: Desafios e Boas Práticas

Levar os conceitos de regressão do papel para o mundo real envolve mais do que apenas aplicar fórmulas. A prática da modelagem preditiva é um ciclo iterativo que exige atenção a detalhes e uma compreensão profunda dos dados e do problema de negócio. Um dos primeiros desafios é a **preparação dos dados**: dados brutos raramente estão prontos para serem usados. Limpeza, tratamento de valores ausentes, normalização e codificação de variáveis categóricas são passos cruciais.



Preparação dos Dados

Limpeza, tratamento de valores ausentes, normalização e codificação de variáveis categóricas.



Engenharia de Features

Criar novas features a partir das existentes para melhorar o desempenho do modelo.



Atenção a Viés e Ética

Garantir que o modelo não perpetue ou amplifique desigualdades existentes.



Avaliação Contextual

Entender o contexto e as implicações das previsões além das métricas.

Outro ponto vital é a **engenharia de features**. Muitas vezes, as variáveis originais não são as mais informativas. Criar novas *features* a partir das existentes (como a razão entre duas variáveis ou termos polinomiais) pode melhorar drasticamente o desempenho do modelo. Além disso, é fundamental estar atento a questões éticas e de viés nos dados, garantindo que o modelo não perpetue ou amplifique desigualdades existentes.

Quando usar Regressão Linear?

Para prever valores contínuos como preços, temperaturas, vendas, etc.

Quando usar Regressão Logística?

Para problemas de classificação binária como Sim/Não, Spam/Não-spam, etc.

A escolha entre Regressão Linear e Logística, por exemplo, depende diretamente do tipo de problema: previsão de valor contínuo ou classificação. A avaliação do modelo não se encerra com as métricas; é preciso entender o contexto e as implicações das previsões. Essa jornada de aprendizado e aplicação contínua é o que transforma um bom modelo em uma solução de impacto.

Consolidação do Conhecimento

Recapitulando a Jornada

Nesta aula, desvendamos os fundamentos da modelagem preditiva, começando pela Regressão Linear, que nos permite prever valores contínuos e entender a influência de múltiplas variáveis. Exploramos como interpretar os coeficientes e como avaliar a performance do modelo através de métricas como R^2 , MSE e MAE. Em seguida, mergulhamos na Regressão Logística, uma ferramenta poderosa para problemas de classificação, compreendendo a função sigmoide, o conceito de log-odds e a crucial matriz de confusão com suas métricas derivadas. Finalizamos com um olhar sobre as tendências de AutoML e XAI, que moldam o futuro da construção e interpretabilidade dos modelos.



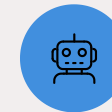
Regressão Linear

- Simples e Múltipla
- Interpretação de coeficientes
- Métricas: R^2 , MSE, MAE



Regressão Logística

- Função sigmoide
- Log-odds
- Matriz de confusão



Tendências

- AutoML
- XAI
- Democratização do ML



Em prática: A capacidade de escolher o modelo de regressão correto, interpretar seus resultados e avaliar sua eficácia é uma habilidade central para qualquer profissional de dados. Use esses conhecimentos para analisar dados de vendas, prever riscos ou classificar comportamentos, sempre buscando entender o "porquê" por trás das previsões.

Autoavaliação

Teste Seus Conhecimentos

1

Qual das seguintes métricas é mais sensível a erros grandes em um modelo de Regressão Linear?

1. R-quadrado (R^2)
2. Erro Absoluto Médio (MAE)
3. Erro Quadrático Médio (MSE)
4. Acurácia

2

A Regressão Logística é utilizada principalmente para qual tipo de problema?

1. Previsão de valores contínuos
2. Agrupamento de dados
3. Classificação de categorias
4. Redução de dimensionalidade

3

Em uma Matriz de Confusão, o que representa um "Falso Positivo"?

1. O modelo previu positivo e o valor real era positivo.
2. O modelo previu negativo e o valor real era negativo.
3. O modelo previu positivo, mas o valor real era negativo.
4. O modelo previu negativo, mas o valor real era positivo.

4

Qual o principal objetivo da Inteligência Artificial Explicável (XAI)?

1. Aumentar a velocidade de treinamento dos modelos.
2. Automatizar a seleção de algoritmos de Machine Learning.
3. Tornar os modelos de IA mais transparentes e compreensíveis.
4. Reduzir o número de variáveis em um dataset.

Questão Discursiva

Explique a diferença fundamental entre a Regressão Linear e a Regressão Logística, abordando o tipo de problema que cada uma resolve e como a função sigmoide é essencial para a Regressão Logística.

Gabarito

Respostas Corretas

Questão 1

c) Erro Quadrático Médio (MSE)

O MSE eleva os erros ao quadrado, penalizando erros grandes mais severamente.

Questão 2

c) Classificação de categorias

A Regressão Logística é um modelo de classificação, não de regressão.

Questão 3

c) O modelo previu positivo, mas o valor real era negativo.

Falso Positivo é um erro do tipo I.



Questão 4

c) Tornar os modelos de IA mais transparentes e compreensíveis.

XAI foca na interpretabilidade e transparência das decisões da IA.

Próximos Passos

Continue Sua Jornada

  **Próxima Aula:** Na Aula 4 – Overfitting, Underfitting e a Validação Cruzada, aprofundaremos nos desafios comuns da modelagem, como evitar que seu modelo aprenda demais ou de menos, e como garantir que ele generalize bem para novos dados.

Recursos Adicionais

Artigo sobre Regressão Linear

Para aprofundar nos pressupostos e aplicações.

Documentação da Scikit-learn (Regressão Logística)

Para exemplos práticos de implementação em Python.

Vídeos explicativos sobre XAI (SHAP e LIME)

Para visualizar como essas técnicas funcionam na prática.

NOTA IMPORTANTE: As informações regulatórias/legais/técnicas desta aula estão atualizadas até 2025. Consulte sempre fontes oficiais para verificar alterações.