MATEMÁTICA APLICADA À INTELIGÊNCIA ARTIFICIAL

Entenda a matemática por trás da IA.

Uma trilha gratuita, em português, que começa do zero e conecta cada conceito a sistemas reais de inteligência artificial.

Sem pré-requisito técnico. Cerca de 6 horas de vídeos verificados, exercícios manuais e gabaritos.

VISÃO GERAL

Quatro fundamentos. Uma aplicação integradora.

Álgebra linear representa dados. Cálculo mede mudança. Probabilidade e estatística tratam incerteza e evidência. Otimização ajusta parâmetros. Deep learning reúne tudo isso em modelos com muitas camadas.

1

Alfabetização

Números, funções, tabelas, gráficos e notação.

2

Representação

Vetores, matrizes, probabilidade e amostras.

3

Aprendizagem

Derivadas, perda, gradientes e otimização.

4

Sistema completo

Avaliação, deep learning, monitoramento e responsabilidade.

Como estudar: leia a explicação, refaça o exemplo no papel, tente a atividade sem consultar o gabarito e só depois assista aos vídeos. Avance quando conseguir explicar o conceito com suas próprias palavras.

00

ETAPA ZERO

Comece sem medo de matemática

Para uma pessoa totalmente leiga, o primeiro objetivo não é decorar fórmulas. É aprender a ler números, símbolos, tabelas e gráficos como formas diferentes de contar a mesma história.

Ler
frações, porcentagens, potências, raízes e sinais

Traduzir
uma situação em tabela, gráfico, função e frase

Conferir
unidades, ordem de grandeza e resultado plausível

O alfabeto que vem antes da IA

Uma porcentagem é uma fração com denominador 100. Dizer que um classificador acertou 92% significa que, em cada 100 casos de um conjunto específico, ele acertou aproximadamente 92. Isso não diz quais oito casos falharam nem se esses erros eram graves.

Números negativos ajudam a representar direção e comparação. Um peso negativo pode reduzir uma pontuação; uma mudança de **-0,3** aponta para o sentido oposto de uma mudança de **+0,3**. Potências aparecem quando medimos áreas, variâncias e erros quadráticos. Raízes desfazem potências. Logaritmos transformam multiplicações em somas e aparecem em funções de perda, mas você só precisa começar entendendo que são a operação inversa da potência.

Função significa regra de transformação

Uma função recebe uma entrada e devolve uma saída. Na regra **y = 2x + 1**, o valor de *x* entra, é multiplicado por dois e recebe mais um. Quando x vale 0, 1, 2 e 3, y vale 1, 3, 5 e 7. A tabela, o gráfico, a frase e a fórmula descrevem a mesma relação.

```
y = 2x + 1
```

Em IA, o modelo inteiro também é uma função: dados entram, uma previsão sai. Treinar é ajustar os números internos dessa função.

Notação que você encontrará

x_i	O item de posição i. Se x é uma lista de mensagens, x_3 é a terceira.
Σ	Somatório. É uma forma curta de pedir "some todos os termos indicados".
$\approx$	Aproximadamente igual. Útil quando arredondamos uma medida.
$<, >, \leq, \geq$	Menor, maior, menor ou igual, maior ou igual. São usados em limites e decisões.

EXEMPLO RESOLVIDO

Porcentagem não é contexto completo

Em 1.000 mensagens, 70 são spam. Um sistema detecta 63 delas.

1. Proporção de spam: 70 / 1.000 = 0,07.
2. Porcentagem de spam: 0,07 × 100 = 7%.
3. Spam detectado: 63 / 70 = 0,90 = 90%.

```
63 ÷ 70 = 0,90 = 90%
```

O sistema encontrou 90% do spam, mas ainda não sabemos quantas mensagens legítimas ele marcou por engano. Uma métrica isolada nunca conta toda a história.

PRÁTICA GUIADA

Faça no papel

1. Complete a tabela de y = 2x + 1 para x igual a -2, -1, 0, 1 e 2.
2. Marque os cinco pontos em um plano cartesiano.
3. Explique com palavras o que muda em y quando x aumenta uma unidade.
4. Calcule 12% de 250 sem calculadora e confira por uma segunda estratégia.

Gabarito e critério

Na função, os valores de y são -3, -1, 1, 3 e 5. Cada aumento de uma unidade em x aumenta y em duas. Doze por cento de 250 é 30: 10% são 25 e 2% são 5.

Teste rápido: Um modelo acerta 95% dos casos. O que podemos concluir sem outras informações?

AULAS EM PORTUGUÊS

Assista depois de ler

Os vídeos complementam a explicação. Cada cartão informa o nível ou a ressalva pedagógica. Se o player estiver indisponível, use o link direto.

GIS COM GIZ MATHEMATICS 17:09

PORCENTAGEM | COMO CALCULAR PORCENTAGEM | \Prof. Gis/ #01

Preparação. Use para revisar porcentagens antes de interpretar métricas de um modelo.

Abrir no YouTube (https://www.youtube.com/watch?v=OjOyNmTt7Mw)

PRINCIPIA MATEMÁTICA 18:02

O que é uma Função? Entenda tudo em 18 minutos

Núcleo. Consolida a ideia de função como transformação de entrada em saída.

Abrir no YouTube (https://www.youtube.com/watch?v=6OkI2V-kUo4)

DICASDEMAT SANDRO CURIÓ 14:04

Matemática | Gráficos e Tabelas para o ENEM | Semana da matemática

Preparação. Treina leitura crítica de tabelas e gráficos, habilidade usada em todo experimento de IA.

Abrir no YouTube (https://www.youtube.com/watch?v=XzZGAwfKs_k)

01

FUNDAMENTO 1

Álgebra linear

Representar e transformar informação. Vetores organizam características; matrizes combinam essas características; dimensões dizem quais operações fazem sentido.

Representar
um exemplo como vetor de características

Transformar
um vetor por uma matriz de pesos

Comparar
distância, norma e alinhamento sem confundir com significado

Da ficha de características à camada neural

Um **escalar** é um único número. Um **vetor** é uma lista ordenada. Uma **matriz** é uma grade de números. Um **tensor** generaliza essa organização para mais dimensões. A forma, ou *shape*, evita operações impossíveis.

Se uma mensagem é descrita por três características, podemos usar x = [2, 1, 3]. Isso pode significar duas palavras em maiúsculas, um link e três termos de urgência. O vetor não é a mensagem; é uma representação escolhida.

Produto interno	Multiplica componentes correspondentes e soma. Produz uma pontuação ponderada ou uma medida de alinhamento.
Matriz × vetor	Cada linha da matriz cria uma nova combinação das características de entrada.
Norma e distância	Medem tamanho e separação. Escalas diferentes podem distorcer comparações.
Transformação afim	z = Wx + b. É a operação central de uma camada neural antes da ativação.
Combinação linear	Soma de vetores multiplicados por pesos. Mostra quais saídas podemos construir a partir das entradas.

Segunda passagem

Base, independência linear, posto, projeções, autovetores e decomposição em valores singulares são importantes, mas não precisam bloquear seu início. Eles ajudam a entender redução de dimensionalidade, estabilidade e direções dominantes dos dados.

Cuidado: proximidade entre embeddings não garante igualdade de significado. O resultado depende do modelo, dos dados, da normalização e da tarefa.

EXEMPLO RESOLVIDO

Produto interno como pontuação

Use x = [2, 1, 3] e pesos w = [1, 2, -1].

$$w^T x = 1\times2 + 2\times1 - 1\times3 = 1$$

O primeiro sinal adiciona dois pontos; o link adiciona dois; a terceira característica remove três. O escore final é 1.

Agora use uma matriz com duas linhas:

```
W = [[1,0,−1],[0,2,1]]
Wx = [−1,5]
```

A matriz recebe três números e devolve dois. Conferir as dimensões evita a maioria dos erros iniciais.

PRÁTICA GUIADA

Cheque dimensões antes de calcular

1. Identifique a dimensão de x = [4, 2, 1].
2. Diga a forma de uma matriz com duas saídas e três entradas.
3. Calcule o produto interno entre [1, 0, 2] e [3, 4, -1].
4. Explique por que uma matriz 2 por 3 não multiplica diretamente um vetor de dimensão 4.

Gabarito e critério

x tem dimensão 3. A matriz deve ser 2 por 3. O produto interno é 1. A operação com dimensão 4 falha porque cada linha tem três pesos, mas receberia quatro componentes.

Teste rápido: Uma matriz W de forma 4 por 3 recebe um vetor x de dimensão 3. Qual é a dimensão de Wx?

AULAS EM PORTUGUÊS

Assista depois de ler

Os vídeos complementam a explicação. Cada cartão informa o nível ou a ressalva pedagógica. Se o player estiver indisponível, use o link direto.

TODA MATÉRIA 7:18

O que são VETORES?

Núcleo. Primeira exposição visual a vetores; o portal faz a ponte para listas de características.

Abrir no YouTube (https://www.youtube.com/watch?v=VKVczB6311I)

KHAN ACADEMY BRASIL 10:45

Produto escalar

Núcleo. Mostra o produto escalar que aparece em pontuações, similaridade e atenção.

Abrir no YouTube (https://www.youtube.com/watch?v=OtGMHiOmMGE)

DICASDEMAT SANDRO CURIÓ 10:59

MATRIZES | DESTRAVE EM 10 MINUTOS

Preparação. Apresenta notação e dimensões de matrizes, sem ainda cobrir toda a álgebra linear.

Abrir no YouTube (https://www.youtube.com/watch?v=HNpSOOhtqkA)

KHAN ACADEMY BRASIL 5:30

A multiplicação de uma matriz por uma matriz

Núcleo. Explica a lógica da multiplicação de matrizes; pause para conferir cada dimensão.

Abrir no YouTube (https://www.youtube.com/watch?v=zhbuYu8w1z0)

DATASTAX DEVELOPERS 12:24

Embeddings e Vector Search para IA Generativa - O que são, como gerar e usar | Samuel Matioli

Intermediário. Mostra embeddings e busca vetorial com código e uma ferramenta específica.

Abrir no YouTube (https://www.youtube.com/watch?v=ygRurMXa5uw)

02

FUNDAMENTO 2A

Probabilidade

Raciocinar sob incerteza. Probabilidade ajuda a falar sobre eventos, evidências e previsões sem transformar possibilidade em certeza.

Distinguir
probabilidade conjunta, marginal e condicional

Atualizar
uma hipótese usando taxa-base e evidência

Interpretar
esperança, variância e distribuição

O universo muda quando há uma condição

Um evento é um conjunto de resultados. Probabilidade condicional significa restringir o universo aos casos que satisfazem a condição. P(spam | alerta) pergunta: entre as mensagens que receberam alerta, quantas eram spam? Isso é diferente de P(alerta | spam): entre as mensagens realmente spam, quantas receberam alerta?

Conjunta	Chance de dois eventos ocorrerem juntos.
Marginal	Chance de um evento, somando ou ignorando outras categorias.
Condicional	Chance dentro de um subconjunto já filtrado.
Independência	Saber que um evento ocorreu não muda a probabilidade do outro, dentro do modelo adotado.
Esperança	Média de longo prazo de muitas repetições, não necessariamente um resultado possível.

Bayes em uma frase

Comece com um peso inicial para cada hipótese. Observe uma evidência. Reavalie os pesos conforme a compatibilidade da evidência com cada hipótese. A taxa-base importa: um evento raro pode continuar raro mesmo depois de um teste aparentemente bom.

Aleatório não significa uniforme. Um processo pode ser aleatório e ainda favorecer alguns resultados.

EXEMPLO RESOLVIDO

Alerta de fraude

Em 10.000 compras, 100 são fraude. O sistema detecta 90 fraudes, mas também alerta 495 compras legítimas.

1. Alertas corretos: 90.
2. Total de alertas: 90 + 495 = 585.
3. Fraudes entre alertas: 90 / 585 ≈ 15,4%.

```
P(fraude | alerta) ≈ 15,4%
```

O sistema detectou 90% das fraudes, mas a maioria dos alertas é falsa porque fraude é rara. Isso não torna o detector inútil; mostra que a decisão precisa considerar custo, revisão humana e taxa-base.

PRÁTICA GUIADA

Troque a direção da pergunta

1. Escreva em palavras P(alerta | fraude).
2. Escreva em palavras P(fraude | alerta).
3. Explique por que as duas probabilidades não são iguais.
4. Imagine que fraude passe de 1% para 10%. Sem mudar o detector, o que tende a acontecer com a proporção de alertas verdadeiros?

Gabarito e critério

A primeira pergunta mede detecção entre fraudes; a segunda mede credibilidade entre alertas. Elas usam denominadores diferentes. Se a taxa-base de fraude sobe, a proporção de alertas verdadeiros tende a subir.

Teste rápido: Qual pergunta mede a chance de uma mensagem ser spam depois que o modelo gerou um alerta?

AULAS EM PORTUGUÊS

Assista depois de ler

Os vídeos complementam a explicação. Cada cartão informa o nível ou a ressalva pedagógica. Se o player estiver indisponível, use o link direto.

DICASDEMAT SANDRO CURIÓ 13:02

PROBABILIDADE | APRENDA EM 13MIN

Preparação. Introdução por contagem de casos, antes de probabilidade condicional e Bayes.

Abrir no YouTube (https://www.youtube.com/watch?v=iNCkGogNtKI)

DIDÁTICA TECH 20:29

Entenda o Teorema de Bayes (ótima explicação!)

Núcleo. Excelente ponte para taxa-base e atualização de crenças; os números do exemplo são didáticos.

Abrir no YouTube (https://www.youtube.com/watch?v=I643PqSrETM)

03

FUNDAMENTO 2B

Cálculo

Descrever mudança e sensibilidade. Derivadas mostram como uma saída reage a pequenas mudanças; gradientes juntam essas sensibilidades para muitos parâmetros.

Interpretar
derivada como taxa de variação local

Ler
derivadas parciais e gradiente

Conectar
regra da cadeia à retropropagação

Da inclinação ao ajuste de milhões de parâmetros

Uma derivada não é só uma regra algébrica. Ela responde: se eu mudar um pouco a entrada, quanto a saída tende a mudar aqui? O velocímetro mostra uma taxa instantânea; o odômetro acumula distância. Essa é a diferença intuitiva entre derivada e integral.

Quando uma função possui várias entradas, calculamos uma derivada parcial para cada uma. O **gradiente** reúne essas derivadas em um vetor. Ele aponta para a direção local de subida mais rápida. Para reduzir uma perda, normalmente caminhamos no sentido oposto.

Limite	Descreve o comportamento de uma função quando a entrada se aproxima de um ponto.
Derivada	Taxa de variação instantânea e inclinação local.
Gradiente	Vetor de sensibilidades de uma função em relação a várias entradas.
Regra da cadeia	Combina efeitos locais quando uma transformação alimenta outra.
Aproximação local	Usa inclinação e curvatura próximas para estimar como a função se comporta.

O que a retropropagação faz

Uma rede é uma composição de funções. A retropropagação aplica a regra da cadeia da saída para a entrada e calcula o quanto cada parâmetro influenciou a perda. Ela **calcula gradientes**. O otimizador usa esses gradientes para decidir a atualização.

Gradiente zero não garante mínimo. Pode ser máximo ou ponto de sela. O gradiente também é informação local, não um mapa completo.

EXEMPLO RESOLVIDO

Perda quadrática simples

Considere $L(w) = (w - 3)^2$. A derivada é $L'(w) = 2(w - 3)$.

```
L'(w) = 2(w − 3)
```

1. Em w = 1, L = 4 e L' = -4. Aumentar w reduz a perda localmente.
2. Em w = 3, L = 0 e L' = 0. Estamos no mínimo deste exemplo.
3. Em w = 4, L = 1 e L' = 2. Diminuir w reduz a perda localmente.

O sinal mostra a direção da subida. O sentido negativo do gradiente aponta a descida local.

PRÁTICA GUIADA

Construa a intuição antes da regra

1. Faça uma tabela de L(w) para w igual a 0, 1, 2, 3, 4, 5 e 6.
2. Desenhe aproximadamente a curva.
3. Marque onde a inclinação é negativa, zero e positiva.
4. Explique por que seguir o gradiente aumentaria L, enquanto seguir o gradiente negativo tende a diminuir L.

Gabarito e critério

As perdas são 9, 4, 1, 0, 1, 4 e 9. À esquerda de 3 a inclinação é negativa; em 3 é zero; à direita é positiva.

Teste rápido: Em uma rede neural, qual componente usa a regra da cadeia para calcular sensibilidades?

AULAS EM PORTUGUÊS

Assista depois de ler

Os vídeos complementam a explicação. Cada cartão informa o nível ou a ressalva pedagógica. Se o player estiver indisponível, use o link direto.

PROFESSOR JULIO LOMBALDO 14:59

Mas afinal, o que é a Derivada?!

Núcleo. Constrói a derivada como inclinação e taxa instantânea de mudança.

Abrir no YouTube (https://www.youtube.com/watch?v=4kjaMznGUnY)

TEM CIÊNCIA 12:43

LIMITE: a Ideia Fundamental do Cálculo [LIMITES]

Opcional. A parte intuitiva é útil; a formalização rigorosa vai além do necessário na primeira passagem.

Abrir no YouTube (https://www.youtube.com/watch?v=8jaLYCGG7io)

04

FUNDAMENTO 3

Probabilidade e estatística

Aprender com amostras e medir incerteza. Estatística ajuda a separar sinal, ruído, viés e variação, além de avaliar se um modelo generaliza para casos novos.

Separar
população, amostra, treino, validação e teste

Medir
média, dispersão, erro e incerteza

Avaliar
matriz de confusão, precisão, revocação e calibração

O modelo aprende com uma amostra, não com o mundo inteiro

A população é o conjunto sobre o qual queremos concluir. A amostra é o que realmente observamos. Se a amostra não representa a população, uma conta perfeita ainda pode produzir uma conclusão ruim.

Média resume centro; variância e desvio-padrão resumem dispersão. Duas coleções podem ter a mesma média e comportamentos totalmente diferentes. Correlação mostra associação, não prova causalidade. Uma terceira variável pode influenciar as duas.

Treino	Dados usados para ajustar os parâmetros.
Validação	Dados usados para escolher configuração, limiar e momento de parar.
Teste	Dados preservados para estimar desempenho final em exemplos ainda não usados.
Precisão	Entre os alertas positivos, quantos estavam corretos.
Revocação	Entre os positivos reais, quantos foram encontrados.
Calibração	Quando previsões de 70% se confirmam perto de 70% em muitos casos comparáveis.

Sobreajuste

Um modelo sobreajustado memoriza particularidades do treino e perde desempenho em dados novos. Um modelo subajustado é simples demais ou foi treinado de forma insuficiente. O objetivo é generalizar, não apenas reduzir a perda de treino.

Vazamento de dados: qualquer informação do futuro ou do conjunto de teste que influencia o treinamento produz uma avaliação otimista e enganosa.

EXEMPLO RESOLVIDO

Uma métrica pode melhorar enquanto outra piora

Em 100 mensagens, 20 são spam. O detector encontra 18 e marca incorretamente 8 mensagens legítimas.

1. Verdadeiros positivos: 18.
2. Falsos positivos: 8.
3. Falsos negativos: 2.
4. Verdadeiros negativos: 72.

```
Precisão = 18 / (18 + 8) ≈ 69,2%
```

```
Revocação = 18 / (18 + 2) = 90%
```

Detectar 90% do spam não significa que cada alerta tenha 90% de chance de ser correto.

PRÁTICA GUIADA

Escolha a métrica pelo custo do erro

1. Calcule a acurácia do exemplo de spam.
2. Imagine que um falso positivo bloqueia uma mensagem urgente. Você priorizaria precisão ou revocação? Justifique.
3. Agora imagine triagem de uma doença grave em uma etapa inicial. O que muda?
4. Explique por que o limiar de decisão deve ser escolhido com validação, não com o teste final.

Gabarito e critério

A acurácia é (18 + 72) / 100 = 90%. O peso de precisão e revocação depende do custo de cada erro. O conjunto de teste deve permanecer intocado para não virar parte do ajuste.

Teste rápido: Entre todos os casos realmente positivos, qual métrica mede a proporção encontrada pelo modelo?

AULAS EM PORTUGUÊS

Assista depois de ler

Os vídeos complementam a explicação. Cada cartão informa o nível ou a ressalva pedagógica. Se o player estiver indisponível, use o link direto.

ORIVALDO SANTANA 9:54

Introdução ML: Treinamento, Teste e Validação

Núcleo. Mostra por que treino, validação e teste têm funções diferentes.

Abrir no YouTube (https://www.youtube.com/watch?v=YhN9g32zDR4)

DIDÁTICA TECH 12:34

O que é Overfitting e Underfitting? (Introdução a Machine Learning - Aula 4)

Núcleo. Explica subajuste, sobreajuste e generalização com exemplos visuais.

Abrir no YouTube (https://www.youtube.com/watch?v=IHAb3NHDahU)

IA EXPERT ACADEMY 7:28

Precision e recall

Núcleo. Use junto ao exemplo de spam para separar precisão de revocação.

Abrir no YouTube (https://www.youtube.com/watch?v=fC_C695yVOk)

MEIGAROM | COMUNIDADE DS 15:27

TUDO que você precisa saber sobre Matriz de Confusão!

Núcleo. Organiza verdadeiros e falsos positivos e negativos em uma matriz de confusão.

Abrir no YouTube (https://www.youtube.com/watch?v=FZqMCgCbo3U)

05

FUNDAMENTO 4

Otimização

Escolher parâmetros para atingir um objetivo. A função de perda transforma desempenho em um número; o algoritmo de otimização procura parâmetros que reduzam esse número sem confundir memorização com aprendizagem.

Definir
parâmetros, objetivo, restrições e regularização

Executar
passos simples de descida do gradiente

Diagnosticar
passo grande, passo pequeno e parada

Treinar é procurar, não programar cada resposta

Parâmetros são os números ajustáveis do modelo. A função objetivo mede o que queremos melhorar. Restrições limitam as escolhas. A regularização adiciona uma preferência, como soluções mais simples ou pesos menores.

$$\theta^* = \arg\min_\theta L(\theta)$$

A descida do gradiente atualiza os parâmetros na direção oposta à subida local. A taxa de aprendizagem α controla o tamanho do passo.

$$\theta_{t+1} = \theta_t - \alpha \nabla L(\theta_t)$$

Mínimo local	Melhor ponto em uma vizinhança. Não precisa ser o melhor de todo o espaço.
Convexidade	Geometria que oferece garantias fortes. Redes profundas costumam ser não convexas.
Minibatch	Pequeno grupo de exemplos usado para estimar o gradiente com custo menor e algum ruído.
Regularização	Preferência adicional que ajuda a controlar complexidade e sobreajuste.
Parada antecipada	Interrompe o treino quando a validação deixa de melhorar.

Menor perda de treino não implica melhor modelo futuro. O modelo pode apenas memorizar os exemplos usados para ajustar os parâmetros.

EXEMPLO RESOLVIDO

Descida do gradiente à mão

Use $L(w) = (w - 4)^2$, $L'(w) = 2(w - 4)$ e comece em $w = 0$.

Com $\alpha = 0{,}25$:

1. $w = 0 - 0{,}25(-8) = 2$.
2. $w = 2 - 0{,}25(-4) = 3$.
3. $w = 3 - 0{,}25(-2) = 3{,}5$.

```
Perda: 16 → 4 → 1 → 0,25
```

Com $\alpha = 1$, w alterna entre 0 e 8, mantendo perda 16. Passo maior não significa treino melhor.

PRÁTICA GUIADA

Compare taxas de aprendizagem

1. Repita o exemplo com taxa 0,1 por quatro passos.
2. Repita com taxa 0,5 por dois passos.
3. Descreva o que acontece quando a taxa é pequena demais.
4. Descreva o que pode acontecer quando a taxa é grande demais.

Gabarito e critério

Com taxa pequena, a perda pode cair de forma estável, mas lentamente. Com taxa grande, o processo pode ultrapassar o vale, oscilar ou divergir.

Teste rápido: Qual componente decide o tamanho da atualização usando o gradiente calculado?

AULAS EM PORTUGUÊS

Assista depois de ler

Os vídeos complementam a explicação. Cada cartão informa o nível ou a ressalva pedagógica. Se o player estiver indisponível, use o link direto.

ESTATIDADOS 11:10

BackPropagation - O que é? Como aplicar no R em Redes Neurais

Intermediário. Veja depois de derivadas, gradiente e função de perda; a demonstração usa R.

Abrir no YouTube (https://www.youtube.com/watch?v=r7YIj_n95Po)

06

PONTE PARA SISTEMAS REAIS

Do problema ao modelo monitorado

Um bom algoritmo não conserta um objetivo mal definido, dados ruins ou avaliação vazada. O ciclo de vida de machine learning começa antes do treino e continua depois da publicação.

Enquadrar
classificação, regressão e agrupamento

Construir
baseline, divisão de dados e métrica alinhada ao risco

Monitorar
deriva, falhas, custo e impacto depois do lançamento

Primeiro defina a decisão

Classificação escolhe uma categoria. **Regressão** prevê um número. **Agrupamento** procura estruturas sem rótulos prontos. Aprendizado supervisionado usa exemplos com respostas; não supervisionado procura padrões; reforço aprende por interação e recompensas.

1. **Problema:** qual decisão será apoiada e qual erro é mais caro?
2. **Dados:** de onde vieram, quem ficou de fora e quais rótulos podem estar errados?
3. **Baseline:** uma regra simples que o modelo precisa superar.
4. **Treino:** ajuste de parâmetros somente com os dados permitidos.

5. **Validação:** escolha de configuração, limiar e parada.
6. **Teste:** estimativa final, usada com parcimônia.
7. **Publicação:** integração com limites, fallback e supervisão.
8. **Monitoramento:** qualidade, latência, custo, deriva e impacto por grupo.

Treinamento não é inferência

No treinamento, o sistema ajusta parâmetros usando exemplos. Na inferência, usa os parâmetros já ajustados para produzir uma saída. Um modelo de linguagem costuma prever o próximo token condicionado ao contexto. Isso não garante verdade, intenção ou compreensão humana.

Baseline honesto: se uma regra simples obtém 90% porque 90% dos casos pertencem à mesma classe, um modelo com 91% pode não justificar o custo e o risco adicionais.

EXEMPLO DE PROJETO

Da pergunta à operação

1. Decisão: priorizar revisão de pedidos com risco de atraso.
2. Entrada: rota, transportadora, horário, histórico e clima disponível no momento da previsão.
3. Rótulo: entrega atrasou ou não.
4. Risco: usar uma informação registrada depois da entrega causaria vazamento.
5. Métrica: revocação pode importar para encontrar atrasos; precisão importa para não sobrecarregar a equipe.
6. Monitoramento: novas rotas e transportadoras podem mudar a distribuição dos dados.

O modelo é uma peça da decisão. A operação precisa saber o que fazer com o alerta e como corrigir erros.

PRÁTICA GUIADA

Desenhe um sistema antes de escolher o algoritmo

1. Escolha um problema simples de classificação do seu cotidiano.
2. Defina entrada, saída, rótulo e momento exato da previsão.
3. Crie um baseline sem machine learning.
4. Liste um risco de vazamento e um grupo que pode ser prejudicado por erros.
5. Escolha uma métrica e justifique pelo custo do erro.

Gabarito e critério

Não há resposta única. A aprovação exige que entrada e rótulo existam no momento certo, que o baseline seja comparável e que a métrica derive do custo real dos erros.

Teste rápido: Qual conjunto deve ficar preservado para a estimativa final?

AULAS EM PORTUGUÊS

Assista depois de ler

Os vídeos complementam a explicação. Cada cartão informa o nível ou a ressalva pedagógica. Se o player estiver indisponível, use o link direto.

WLADEMIR PRATES 7:24

Aprendizado de Máquina: Supervisionado e Não Supervisionado

Núcleo. Distingue aprendizado supervisionado e não supervisionado de forma curta.

Abrir no YouTube (https://www.youtube.com/watch?v=ZGyY2goDIJY)

ORIVALDO SANTANA 9:54

Introdução ML: Treinamento, Teste e Validação

Núcleo. Mostra por que treino, validação e teste têm funções diferentes.

Abrir no YouTube (https://www.youtube.com/watch?v=YhN9g32zDR4)

DIDÁTICA TECH 12:34

O que é Overfitting e Underfitting? (Introdução a Machine Learning - Aula 4)

Núcleo. Explica subajuste, sobreajuste e generalização com exemplos visuais.

Abrir no YouTube (https://www.youtube.com/watch?v=IHAb3NHDahU)

07

APLICAÇÃO INTEGRADORA

Deep learning

Deep learning não é um quinto ramo matemático. É uma família de modelos que integra álgebra linear, cálculo, probabilidade, estatística e otimização em composições de muitas camadas.

Percorrer
propagação direta, perda, gradientes e atualização

Explicar
por que a não linearidade é necessária

Distinguir
MLP, CNN e Transformer em nível conceitual

Uma rede é uma função parametrizada

$$\hat{y} = f_\theta(x)$$

Cada camada recebe uma representação, combina seus componentes com pesos e viés, aplica uma função não linear e entrega outra representação. Sem não linearidade, várias camadas lineares seriam equivalentes a uma única transformação linear.

1. **Propagação direta:** calcula a previsão.
2. **Perda:** compara previsão e alvo.
3. **Retropropagação:** calcula gradientes pela regra da cadeia.
4. **Otimizador:** atualiza os parâmetros.
5. **Validação:** verifica se o modelo melhora fora do treino.

MLP	Camadas densas. Bom ponto de partida para entender a mecânica geral.
CNN	Explora estrutura local e compartilhamento de pesos, muito usada em imagens.
Transformer	Usa atenção para combinar informações de diferentes posições; domina muitos sistemas de linguagem.
Logits	Escores antes da conversão em distribuição. Softmax não garante calibração nem verdade.
Época e lote	Época percorre o conjunto de treino; lote é o grupo processado por atualização.

Evite antropomorfizar. Neurônios artificiais são operações matemáticas, não cópias fiéis de células biológicas. Prever tokens não equivale automaticamente a pensar, compreender ou dizer a verdade.

PROPAGAÇÃO DIRETA

Duas unidades ReLU

$$h_1 = \mathrm{ReLU}(x_1 + x_2 - 2)$$

$$h_2 = \mathrm{ReLU}(2x_1 - x_2)$$

$$s = 2h_1 - h_2$$

ReLU(z) = max(0, z). Para x = (1,2), h = (1,0) e s = 2. Para x = (0,2), h = (0,0) e s = 0. Se s > 0 representa classe A, só o primeiro exemplo entra em A.

A álgebra linear combina entradas; a não linearidade cria novas regiões; a perda mediria o erro; cálculo e otimização ajustariam os pesos.

PRÁTICA GUIADA

Mapeie cada fundamento

1. Calcule a rede para x = (2, 0).
2. Troque o limiar de classe de 0 para 2 e observe a decisão.
3. Aponte onde aparece álgebra linear.
4. Explique onde cálculo, probabilidade e otimização entrariam durante o treino.
5. Diga por que empilhar apenas funções lineares não cria o mesmo poder de representação.

Gabarito e critério

Para x = (2,0), h1 = 0, h2 = 4 e s = -4, portanto não A. Pesos e somas usam álgebra linear; gradientes usam cálculo; a perda e avaliação usam estatística; a atualização usa otimização.

Teste rápido: Quem atualiza os pesos depois que os gradientes foram calculados?

AULAS EM PORTUGUÊS

Assista depois de ler

Os vídeos complementam a explicação. Cada cartão informa o nível ou a ressalva pedagógica. Se o player estiver indisponível, use o link direto.

CÓDIGO LOGO 8:11

COMO FUNCIONAM AS REDES NEURAIS?

Núcleo conceitual. Assista lembrando que uma rede artificial não replica literalmente um cérebro.

Abrir no YouTube (https://www.youtube.com/watch?v=Y4lVTmQ43aQ)

ESTATIDADOS 11:10

BackPropagation - O que é? Como aplicar no R em Redes Neurais

Intermediário. Veja depois de derivadas, gradiente e função de perda; a demonstração usa R.

Abrir no YouTube (https://www.youtube.com/watch?v=r7YIj_n95Po)

DATASTAX DEVELOPERS 12:24

Embeddings e Vector Search para IA Generativa - O que são, como gerar e usar | Samuel Matioli

Intermediário. Mostra embeddings e busca vetorial com código e uma ferramenta específica.

Abrir no YouTube (https://www.youtube.com/watch?v=ygRurMXa5uw)

DIDÁTICA TECH 1H 27MIN

Transformers Explicado: Como o ChatGPT Realmente Pensa e Funciona

Avançado. Aula longa sobre Transformers. O título antropomorfiza: prever tokens não é pensar como uma pessoa.

Abrir no YouTube (https://www.youtube.com/watch?v=buc3S1baP34)

08

CAMADA OBRIGATÓRIA

IA responsável

Desempenho médio não basta. Um sistema útil precisa considerar viés, privacidade, segurança, transparência, direitos, supervisão humana e impactos diferentes entre grupos.

Investigar
origem, cobertura e limitações dos dados

Comparar
erros e impactos por grupos relevantes

Governar
limites, revisão humana, documentação e contestação

A matemática mede o que escolhemos medir

Uma função de perda transforma objetivos humanos em números. Essa tradução sempre deixa algo de fora. Se os rótulos refletem decisões históricas injustas, o modelo pode aprender e ampliar o padrão. Se alguns grupos aparecem pouco nos dados, a média geral pode esconder falhas graves.

Viés de dados	Amostragem, medição ou rótulo representam grupos e situações de forma desigual.
Equidade	Não existe uma única métrica universal. Critérios diferentes podem entrar em conflito.
Privacidade	Minimizar coleta, controlar acesso, definir retenção e avaliar risco de reidentificação.
Segurança	Tratar abuso, entradas adversariais, vazamento de dados e dependências externas.
Supervisão	Definir quando uma pessoa revisa, corrige, contesta e interrompe o sistema.

Perguntas antes de publicar

1. Quem recebe o benefício e quem assume o risco?
2. Quais pessoas não estão representadas nos dados?
3. Qual erro exige revisão humana obrigatória?
4. Existe rota de contestação e correção?
5. O sistema registra versões, fontes e limitações?
6. Como detectar deriva e retirar o modelo do ar?

Transparência honesta: explique finalidade, dados permitidos, métricas, limitações e responsável. Não prometa neutralidade, objetividade ou ausência de erro.

EXEMPLO DE AUDITORIA

Reconhecimento de voz

O sistema acerta 92% no total. Em um grupo A, acerta 97%; em um grupo B, 72%.

1. A média geral parece alta.
2. O desempenho por grupo revela uma diferença operacional importante.
3. A causa pode envolver amostragem, ruído, sotaque, equipamento ou rótulos.
4. A resposta não é ocultar a diferença nem atribuí-la ao grupo. É investigar dados, processo e uso.

Mesmo após correções, pode ser necessário limitar o uso, oferecer alternativa manual e monitorar continuamente.

PRÁTICA GUIADA

Faça uma revisão de risco

1. Escolha o sistema desenhado no módulo anterior.
2. Liste quem pode ser afetado por falso positivo e falso negativo.
3. Defina duas métricas por grupo, não só uma média global.
4. Escreva uma limitação que deve aparecer para o usuário.
5. Defina quando uma pessoa pode revisar, contestar e desfazer a decisão.

Gabarito e critério

A aprovação exige riscos ligados a pessoas reais, métricas segmentadas pertinentes, limitação explícita e uma rota operacional de revisão e contestação.

Teste rápido: Um modelo tem ótima média geral. Isso basta para declarar o sistema justo?

AULAS EM PORTUGUÊS

Assista depois de ler

Os vídeos complementam a explicação. Cada cartão informa o nível ou a ressalva pedagógica. Se o player estiver indisponível, use o link direto.

KHAN ACADEMY BRASIL 3:27

Ética e IA

Núcleo. Introdução breve aos dilemas éticos de sistemas de IA.

Abrir no YouTube (https://www.youtube.com/watch?v=8LhauBqbiCI)

CANAL USP 22:20

Implicações éticas e sociais da inteligência artificial - Dora Kaufman - USP Talks #48

Aprofundamento. Discussão institucional sobre impactos sociais, poder e responsabilidade.

Abrir no YouTube (https://www.youtube.com/watch?v=KZGYOl5mAzs)

BASE ACADÊMICA

Fontes e limites

As explicações foram organizadas para iniciantes a partir de referências acadêmicas abertas e materiais institucionais. Os vídeos são complementos de terceiros e foram verificados como públicos em 24 de agosto de 2026.

01 **Mathematics for Machine Learning**

Deisenroth, Faisal e Ong. Estrutura aberta de álgebra linear, cálculo, probabilidade e otimização antes das aplicações.
https://mml-book.github.io/

02 **MIT OpenCourseWare: Linear Algebra**

Gilbert Strang. Sistemas, espaços vetoriais, autovalores e matrizes.
https://ocw.mit.edu/courses/18-06-linear-algebra-spring-2010/

03 **MIT OpenCourseWare: Single Variable Calculus**

Derivadas, integrais, aproximação e otimização elementar.
https://ocw.mit.edu/courses/18-01sc-single-variable-calculus-fall-2010/

04 **MIT OpenCourseWare: Multivariable Calculus**

Derivadas parciais, regra da cadeia, gradiente e otimização.

https://ocw.mit.edu/courses/18-02sc-multivariable-calculus-fall-2010/

05 **MIT OpenCourseWare: Introduction to Probability and Statistics**

Variáveis aleatórias, distribuições, Bayes, inferência e regressão.
https://ocw.mit.edu/courses/18-05-introduction-to-probability-and-statistics-spring-2022/

06 **Convex Optimization**

Boyd e Vandenberghe. Referência aberta para uma segunda passagem em otimização.
https://web.stanford.edu/~boyd/cvxbook/

07 **Deep Learning**

Goodfellow, Bengio e Courville. Livro aberto sobre fundamentos, redes, regularização e otimização.
https://www.deeplearningbook.org/

08 **Curso intensivo de Machine Learning do Google**

Material em português sobre dados, modelos, generalização, métricas e IA responsável.
https://developers.google.com/machine-learning/crash-course?hl=pt-br

Limite editorial: a trilha ensina fundamentos e interpretação. Ela não substitui um curso formal completo, não oferece certificação e não transforma métricas em garantias. Ferramentas, interfaces e exemplos de produto podem mudar; os conceitos matemáticos e as referências canônicas são o eixo durável.