Você aprovaria estes quatro clientes?

Crédito pessoal, horizonte de 12 meses

Ana
R$ 7.000 por mês · comprometimento 22% · histórico não · utilização 30% · relacionamento 36 meses

Bruno
R$ 4.500 por mês · comprometimento 38% · histórico sim · utilização 65% · relacionamento 8 meses

Carla
R$ 10.000 por mês · comprometimento 48% · histórico não · utilização 85% · relacionamento 60 meses

Diego
R$ 3.500 por mês · comprometimento 55% · histórico sim · utilização 90% · relacionamento 4 meses

Que informação faria você mudar de ideia?

Aprovar exige estimar risco e definir uma política.

Esquema conceitual; sem dados

COMO USAR

Escolha uma opção por cliente. A seleção fica apenas neste dispositivo e não vira estatística da turma.

O desafio da aula

Estimar PD, comparar modelos e transformar risco em decisão.

O que sabemos no momento da proposta?

Seis características por cliente, com unidade e momento de observação

CLIENTE

7.000	22	36	30	0	Agência
4.500	38	8	65	1	Digital
10.000	48	60	85	0	Digital
3.500	55	4	90	1	Parceiro

Ana

Renda	**R$ 7.000**
Comprometimento	**22%**
Relacionamento	**36 meses**
Utilização	**30%**
Histórico	**não**
Canal	**Agência**

Perfil fictício, elegível, sem desfecho revelado.

DICIONÁRIO

Selecione um cabeçalho para ver unidade, definição e momento de observação da variável.

Cada variável descreve uma dimensão do risco. Nenhuma decide sozinha.

Esquema conceitual; sem dados

Probabilidade de quê, para quem e até quando?

Atraso superior a 90 dias em qualquer momento dos 12 meses após a contratação

Evento: atraso superior a 90 dias em qualquer momento dos 12 meses seguintes à contratação.

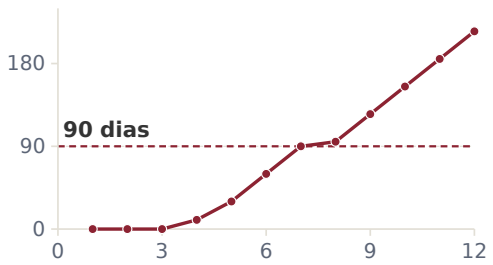
População: solicitantes elegíveis, sem inadimplência corrente na proposta. Unidade: um contrato de crédito pessoal.

Data até a qual conhecemos os dados (mês desde a contratação)

mês 12

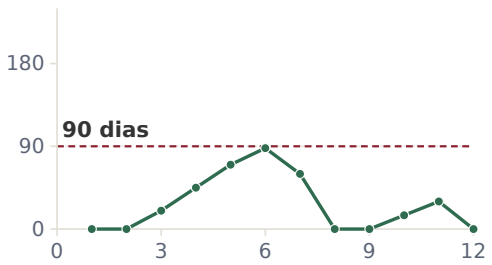
Atingir exatamente 90 dias não satisfaz a definição. O evento exige mais de 90.

CONTRATO A y = 1



evento identificado no mês 8, y = 1

CONTRATO B y = 0



janela completa sem evento, y = 0

CONTRATO C alvo incompleto

janela incompleta, alvo ainda desconhecido

Contratado há 6 meses: a janela de 12 meses ainda não terminou.

Eixo horizontal: meses desde a contratação. Eixo vertical: atraso em dias.

Sem a mesma definição de inadimplência, duas PDs não são comparáveis.

Trajetórias ilustrativas de contratos fictícios; cálculo próprio

Esse dado existia quando o crédito foi aprovado?

Classifique cada informação pela data em que ela ficou disponível

Renda verificada: R$ 4.500 por mês disponibilidade: documento verificado 3 dias antes da proposta	sem classificação
Histórico de atrasos: um atraso de 40 dias disponibilidade: apurado nos 12 meses anteriores à proposta	sem classificação
Utilização de limite: 65% disponibilidade: último extrato fechado antes da proposta	sem classificação
Comprometimento com a parcela proposta: 38% disponibilidade: calculado na simulação da própria proposta	sem classificação
Valor renegociado: R$ 3.200 disponibilidade: registrado 3 meses depois da contratação	sem classificação
Atraso máximo nos 12 meses seguintes: 95 dias disponibilidade: apurado no fim da janela de 12 meses	sem classificação
Consulta de bureau: escore 480 disponibilidade: consultada 2 dias depois da inadimplência	sem classificação

Uma variável excelente depois do evento pode ser inútil para decidir antes dele.

Registros didáticos de disponibilidade; cálculo próprio

A FRONTEIRA DA DECISÃO

Só entra no modelo o que já existia e já estava acessível no momento da proposta.

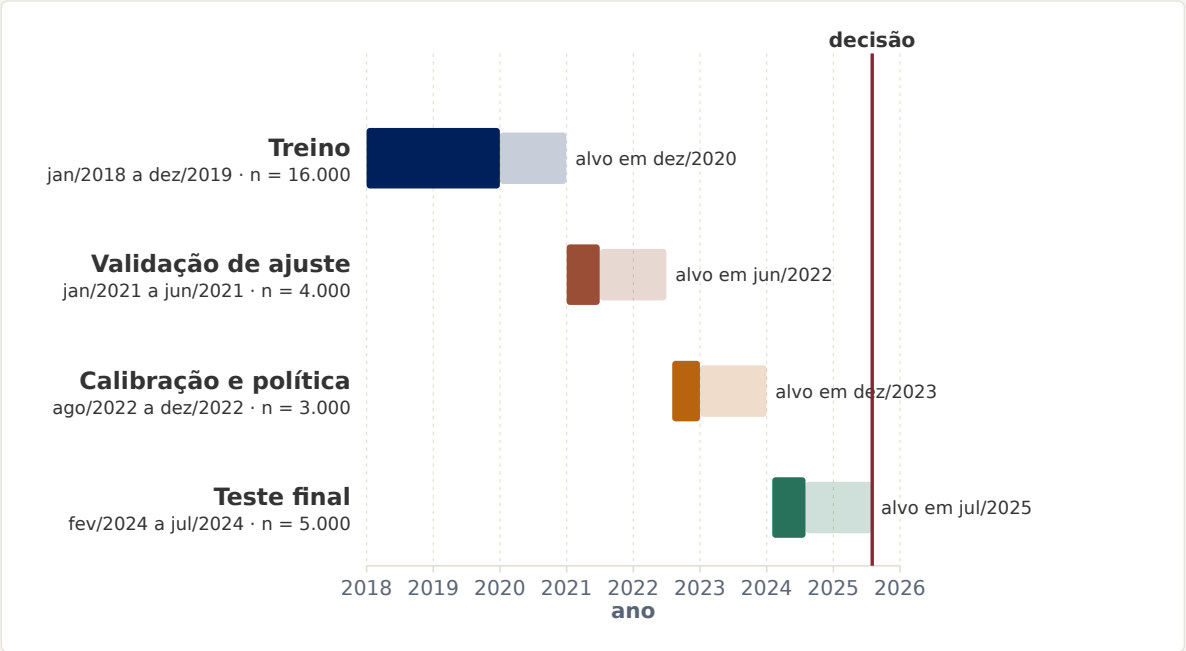
0 de 7 fichas classificadas corretamente.

A REGRA

Vale a data de disponibilidade da informação, não a data do fenômeno que ela descreve.

Aprender no passado, escolher sem olhar o teste

Quatro partições, cada uma com 12 meses de maturação antes de o alvo existir



MOMENTO DA DECISÃO

Com o alvo do teste completo, avaliamos uma única vez as regras já congeladas. O teste não escolhe nada.

Barra sólida: contratações. Barra clara: os 12 meses até o alvo existir. Uma faixa apagada indica alvo ainda desconhecido na data da decisão.

Separar por calendário só funciona se respeitarmos quando cada resultado ficou disponível.

Experimento sintético; semente e versões no notebook

O mesmo cliente, três formas de aprender o risco

A ficha de entrada é a mesma nos três mecanismos

Bruno R$ 4.500 por mês · comprometimento 38% · relacionamento 8 meses · utilização 65% · histórico sim

Ordenar risco: o maior risco fica acima do menor?

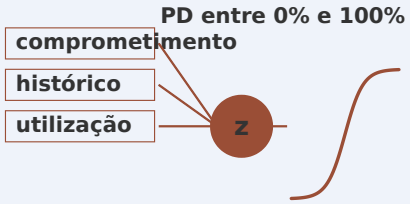
Estimar probabilidades: 10% previsto vira cerca de 10% observado?

Apoiar a decisão: a política resultante melhora o resultado?

Regressão logística

etapa 3 de 3

Combina efeitos em um escore



O escore vira probabilidade

Árvore de decisão

etapa 3 de 3

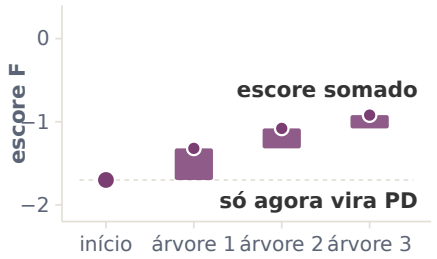
Separa clientes por perguntas

A taxa observada na folha

Gradient boosting

etapa 3 de 3

Acrescenta correções sucessivas ao escore



Só no fim o escore vira probabilidade

Os dados e o objetivo são comuns. Muda a forma de construir a previsão.

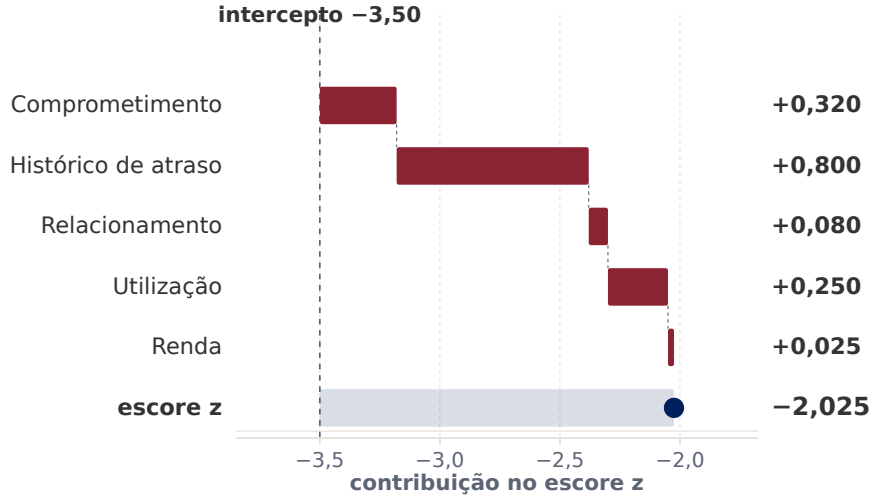
Esquema conceitual; sem dados

Cada informação altera o escore do cliente

Do perfil de referência às diferenças do cliente

$$z = \beta_0 + \sum_j \beta_j (x_j - \bar{x}_j)$$

cada barra abaixo é uma parcela da soma: $\beta_j (x_j - \bar{x}_j)$



Cliente

Perfil de referência: renda R$ 5.000, comprometimento 30%, relacionamento 12 meses, utilização 40% e sem histórico de atraso. O intercepto vale para esse perfil.

CARACTERÍSTICA	CONTA coeficiente × (valor − referência)	CONTRIBUIÇÃO no escore z
Comprometimento (p.p.)	0,04 × (38 − 30)	+0,320
Histórico de atraso (indicador)	0,80 × (1 − 0)	+0,800
Relacionamento (meses)	−0,02 × (8 − 12)	+0,080
Utilização (p.p.)	0,01 × (65 − 40)	+0,250
Renda (R$)	−0,00005 × (4500 − 5000)	+0,025
Intercepto (perfil de referência)	−3,50	

$$z = -2{,}025$$

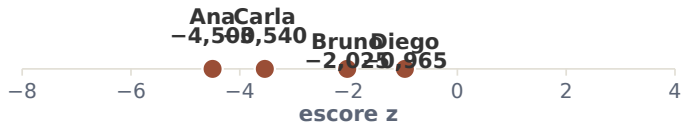
O sinal e o tamanho de cada contribuição dependem do coeficiente e da unidade da variável.

Exemplo didático de logit; cálculo próprio

−2,025 não é uma probabilidade

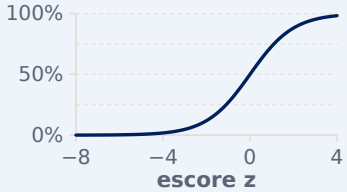
Uma previsão sem restrição pode sair do intervalo entre 0% e 100%

OS ESCORES DOS QUATRO CLIENTES



PRECISAMOS DE UMA TRANSFORMAÇÃO

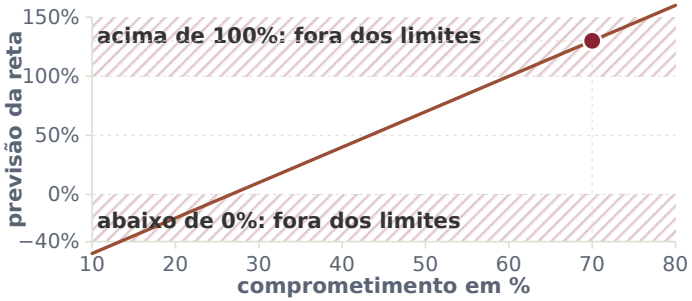
Em vez de restringir a reta, modelamos uma escala livre e a convertemos por uma função que nunca sai do intervalo. É a curva logística do próximo slide.



UMA RETA DIRETO NA PROBABILIDADE

exemplo do problema

$$p_{\text{reta}} = 0{,}10 + 0{,}03\,(comp - 30)$$



Comprometimento

70%

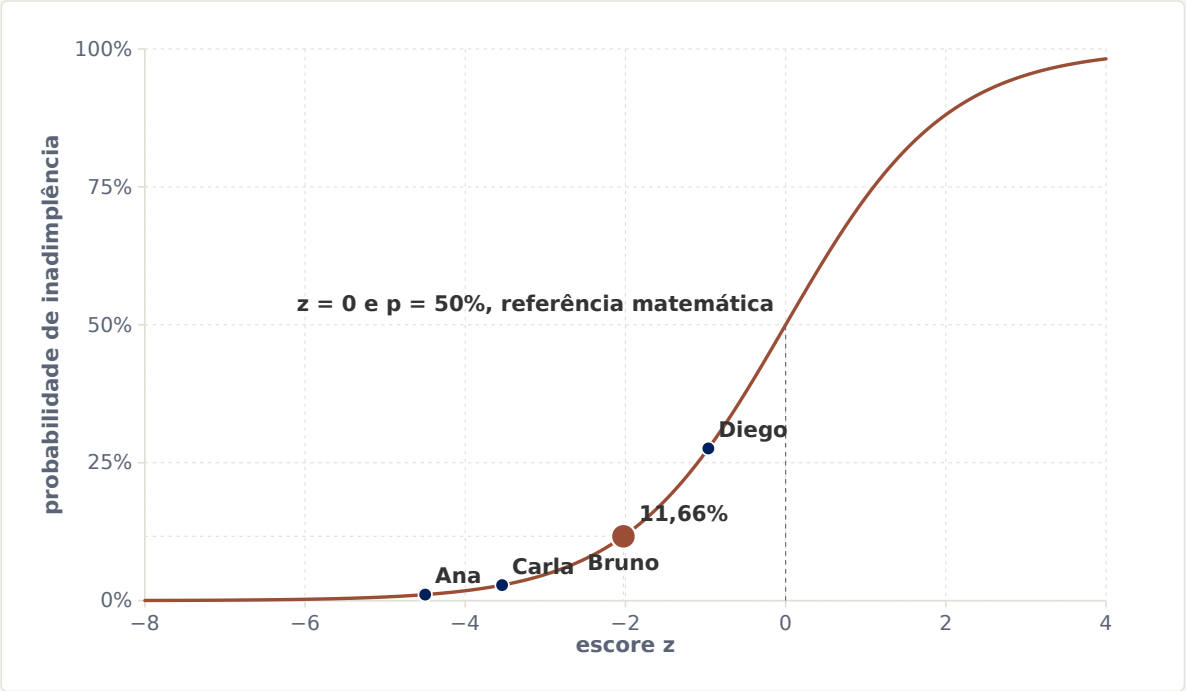
previsão da reta: 130% · fora dos limites de probabilidade

A PD precisa permanecer entre 0% e 100%, qualquer que seja o escore.

Exemplo didático de logit; cálculo próprio

Um escore, uma probabilidade entre 0% e 100%

A curva logística converte qualquer número real em probabilidade



$$p = \frac{1}{1 + e^{-z}}$$

escore z **−2,025**

PD **11,66%**

Leitura: aproximadamente 11,7 eventos em 100 exposições semelhantes, e não uma contagem garantida.

Escore do perfil de Bruno.

Escore z

11,66%

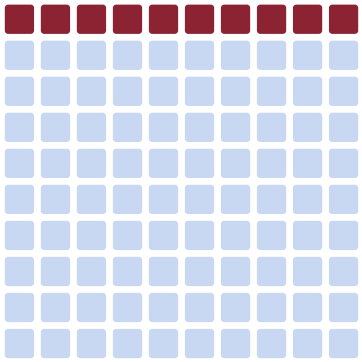
A mesma variação de escore pode gerar mudanças muito diferentes de probabilidade.

Exemplo didático de logit; cálculo próprio

10% de PD significa odds de 1 para 9

Probabilidade, odds e log odds descrevem o mesmo risco em escalas diferentes

FREQUÊNCIA ESPERADA EM 100 EXPOSIÇÕES



10 eventos e 90 não eventos. Representação ilustrativa, não 100 contratos com resultado garantido.

PROBABILIDADE

10 / 100 = 10%

denominador: 100

ODDS

10 / 90 = 0,1111

denominador: 90

LOG ODDS

ln(0,1111) = −2,197

É exatamente a escala em que o logit soma as contribuições: o escore z.

Probabilidade de inadimplência

odds 0,1111

Referências

CONVERSÕES NOS DOIS SENTIDOS

$$\text{odds} = \frac{p}{1-p}$$

$$p = \frac{\text{odds}}{1+\text{odds}}$$

$$z = \ln\left(\frac{p}{1-p}\right)$$

$$p = \frac{1}{1+e^{-z}}$$

Conferência de ida e volta: partindo de 10,00%, a volta por odds devolve 10,00% e a volta por log odds devolve 10,00%.

Nos limites 0% e 100% o log odds vai a infinito. Por isso a escala vai de 1% a 99%.

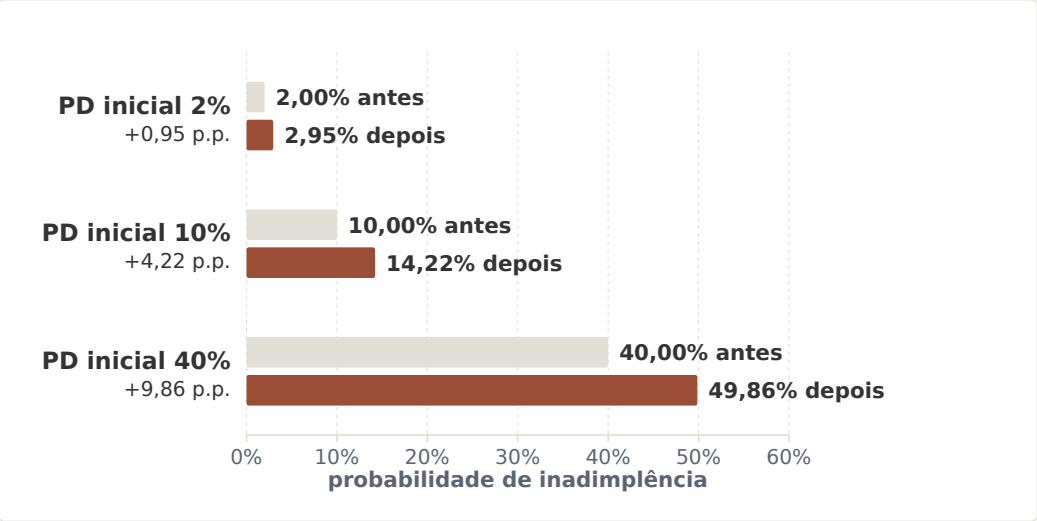
Odds comparam eventos e não eventos. Probabilidade compara eventos e total.

Exemplo didático de logit; cálculo próprio

Mais 10 pontos de comprometimento multiplicam as odds por 1,49

O mesmo multiplicador de odds gera mudanças diferentes de PD

$\Delta comp = +10$ p.p. $\Delta z = 0{,}04 \times 10 = 0{,}4000$ razão de chances = e^0,40 ≈ 1,4918



Uma mudança na variável multiplica as odds. A mudança de PD depende do ponto de partida.

Exemplo didático de logit; cálculo próprio

MUDANÇA APLICADA

Exploração: variação de comprometimento

razão 1,4918

PD INICIAL	ODDS ANTES	ODDS DEPOIS	PD DEPOIS	DIFERENÇA p.p.
2%	0,0204	0,0304	2,95%	+0,95 p.p.
10%	0,1111	0,1658	14,22%	+4,22 p.p.
40%	0,6667	0,9945	49,86%	+9,86 p.p.

Razão de chances de 1,49 não significa risco 49% maior.
Odds e probabilidade são escalas diferentes.

Bruno: como chegamos à PD de 11,66%?

Duas etapas auditáveis: somar o escore e aplicar a logística

FICHA DO CLIENTE

perfil original

Renda em R$ por mês
R$ 4.500

Comprometimento em %
38%

Relacionamento em meses
8 meses

Utilização em %
65%

Histórico de atraso de 15 a 89 dias

Canal: Digital. Informativo: não participa desta fórmula manual.

ETAPA 1: SOMAR O ESCORE

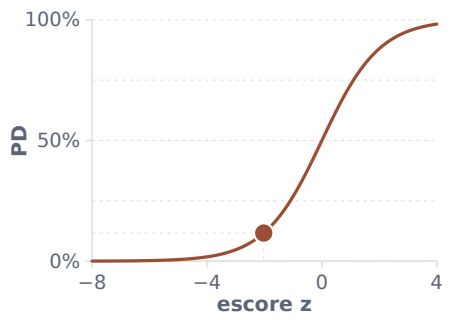
PARCELA	CONTA	VALOR
Intercepto	perfil de referência	−3,500
Comprometimento	0,04 × (38 − 30)	+0,320
Histórico de atraso	0,80 × (1 − 0)	+0,800
Relacionamento	−0,02 × (8 − 12)	+0,080
Utilização	0,01 × (65 − 40)	+0,250
Renda	−0,00005 × (4500 − 5000)	+0,025
Escore z	**soma das parcelas**	**−2,025**

ETAPA 2: APLICAR A LOGÍSTICA

$$p = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{2{,}025}}$$

11,66%

valor sem arredondamento: 0,1166029692



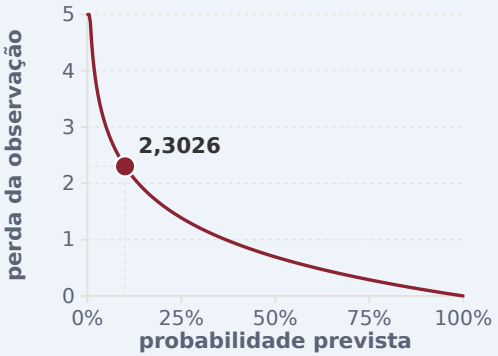
O cálculo tem duas etapas auditáveis: somar o escore e aplicar a logística.

Exemplo didático de logit; cálculo próprio

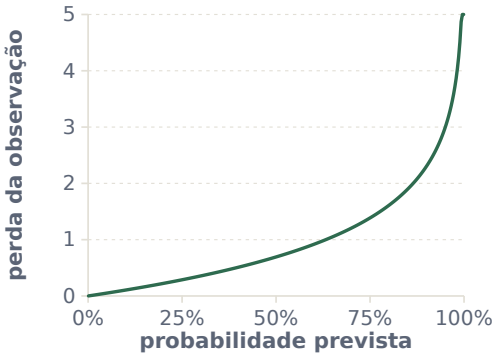
O modelo aprende penalizando previsões incompatíveis com os dados

Perda logarítmica de uma observação, em função da probabilidade prevista

OBSERVAÇÃO COM Y = 1 (HOUVE EVENTO)



OBSERVAÇÃO COM Y = 0 (NÃO HOUVE EVENTO)



$$L = -\left[\, y \ln p + (1-y) \ln(1-p) \,\right]$$

observação **y = 1**
previsão **10,0%**
perda **2,3026**

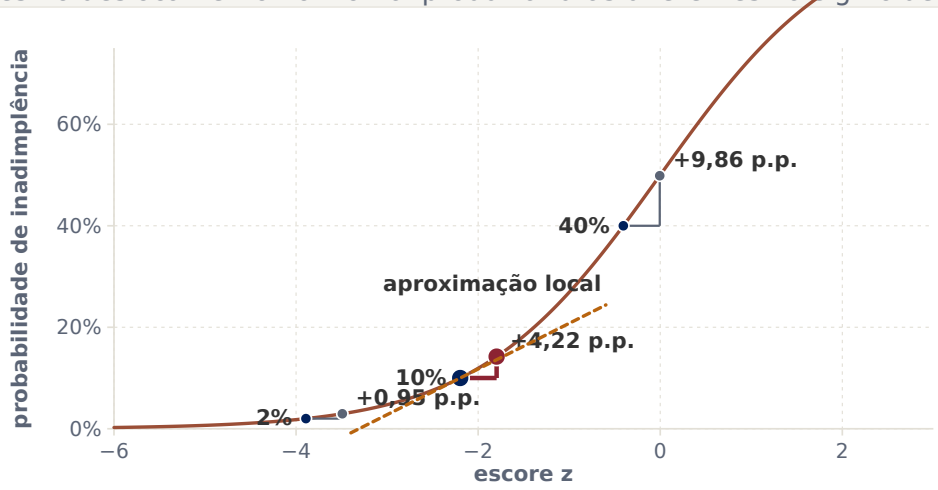
Probabilidade prevista
perda 2,3026
Desfecho observado

OBSERVAÇÃO	y	p prevista	L
nº 1	0	5%	0,0513
nº 2	0	10%	0,1054
nº 3	0	20%	0,2231
nº 4	0	40%	0,5108
nº 5	1	60%	0,5108
nº 6	1	80%	0,2231
Média			**0,2708**

$$\mathrm{L}(\beta) = \frac{1}{n} \sum_{i=1}^{n} -\left[\, y_i \ln p_i + (1 - y_i) \ln(1-p_i) \,\right]$$

Aumentar o escore em 0,40 não aumenta toda PD pelo mesmo valor

O mesmo deslocamento horizontal produz alturas diferentes na sigmoide



$$\Delta comp = 10 \text{ p.p.} \;\Rightarrow\; \Delta z = 0{,}04 \times 10 = 0{,}400$$

$$p : 10{,}00\% \longrightarrow 14{,}22\% \quad (4{,}22 \text{ p.p.})$$

PD inicial

ou valor livre, de 1% a 60%

+4,22 p.p.

Mudança em comprometimento, em pontos

Δz = 0,400

PD INICIAL	VARIAÇÃO EXATA p.p.	APROXIMAÇÃO LOCAL p.p.	DIFERENÇA p.p.
2%	+0,955 p.p.	+0,784 p.p.	−0,171 p.p.
10%	**+4,219 p.p.**	**+3,600 p.p.**	**−0,619 p.p.**
40%	+9,863 p.p.	+9,600 p.p.	−0,263 p.p.

$\frac{\partial p}{\partial\, comp} = \beta\, p\,(1-p) = 0{,}04\, p\,(1-p)$ em probabilidade por ponto percentual; vale para mudanças pequenas.

A resposta em probabilidade é mais intensa na região central da curva.

Exemplo didático de logit; cálculo próprio

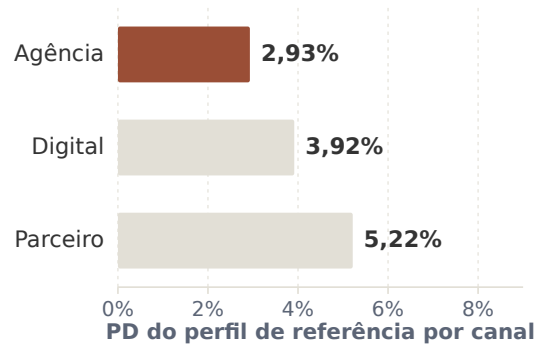
Canal digital não significa duas vezes agência

Categorias entram por comparação com uma referência

REFERÊNCIA ATUAL: AGÊNCIA extensão pedagógica da fórmula

$$z = -3{,}50 + 0{,}30\ 1[\text{digital}] + 0{,}60\ 1[\text{parceiro}]$$

CANAL	INDICADOR DIGITAL	INDICADOR PARCEIRO	CONTRIBUIÇÃO no escore	ESCORE Z	PD
Agência	**0**	**0**	**0,00**	**−3,50**	**2,93%**
Digital	1	0	+0,30	−3,20	3,92%
Parceiro	0	1	+0,60	−2,90	5,22%



O QUE 1, 2 E 3 IMPORIA

Usar canal = 1, 2 e 3 em uma única variável imporia que a diferença entre agência e digital é igual à diferença entre digital e parceiro, e que a ordem econômica existe. Nada nos dados garante isso.

Canal do cliente

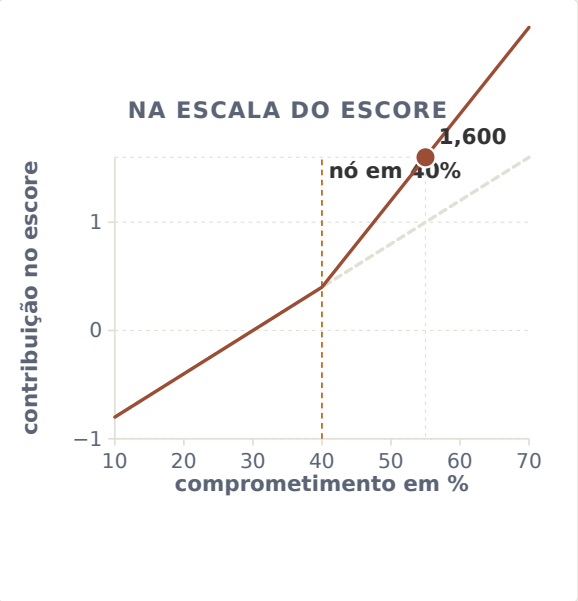
Trocar a referência muda o intercepto e os coeficientes. As três PDs permanecem 2,93%, 3,92% e 5,22% em qualquer parametrização.

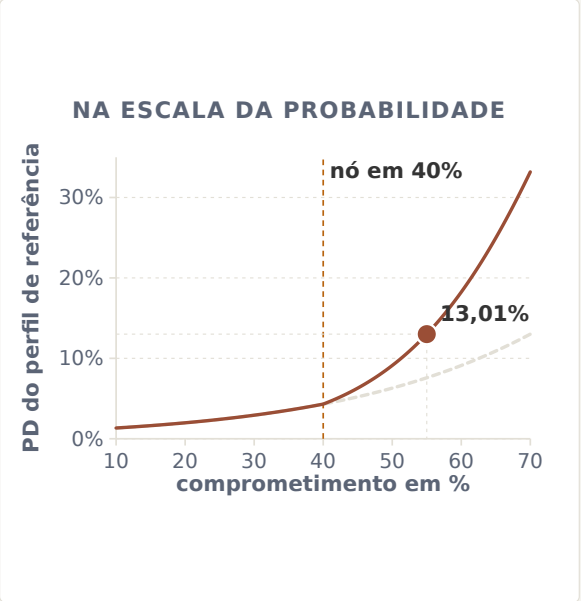
O número atribuído a uma categoria não deve inventar uma ordem econômica.

Extensão pedagógica da fórmula manual; cálculo próprio

O efeito pode se intensificar depois de um limite

Linearidade é nos parâmetros, não na forma das variáveis





$$z = -3{,}50 + 0{,}04\,(comp - 30) + 0{,}04\,\max(comp - 40,\,0)$$

Comprometimento
13,01%

INCLINAÇÃO POR FAIXA

FAIXA	INCLINAÇÃO NO ESCORE por ponto
comprometimento até 40%	0,04
comprometimento acima de 40%	0,08

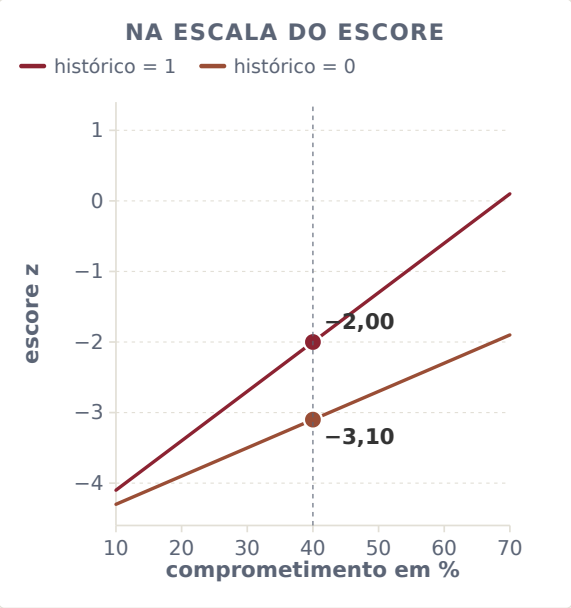
No nó, as duas formas coincidem: contribuição 0,400 em ambos os casos. A função é contínua.

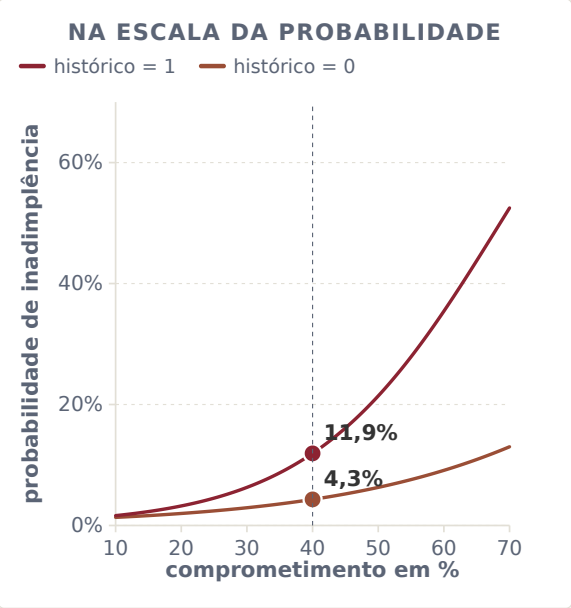
A linearidade é nos parâmetros. A forma das variáveis pode ser flexível.

Cenário separado da fórmula manual; cálculo próprio

O comprometimento pesa igual para quem tem histórico de atraso?

Interação é diferença de inclinação na escala do escore





$$z = -3{,}50 + 0{,}80\,hist + (0{,}04 + \delta\,hist)\,(comp - 30)$$

Comprometimento
termo 0,300

EM COMPROMETIMENTO DE 40%

inclinações no escore	**0,04 e 0,07 por ponto**
diferença no escore	**1,100**
diferença em PD	**+7,61 p.p.**

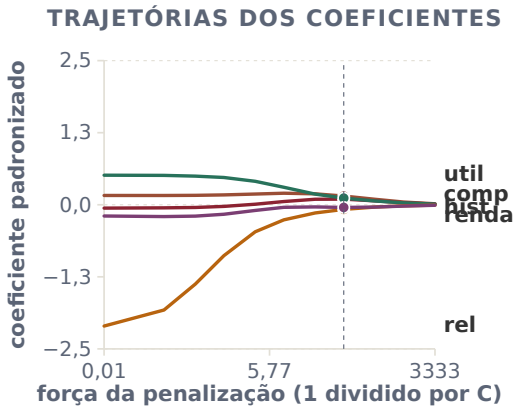
Mesmo sem interação, a diferença em PD varia com o comprometimento: a sigmoide é não linear.

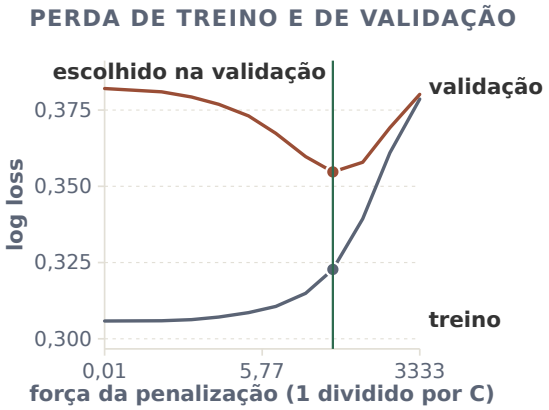
Com interação, o efeito do comprometimento no escore depende do histórico.

Cenário separado da fórmula manual; cálculo próprio

Coeficientes extremos podem estar aprendendo ruído

Trajetórias de coeficientes e perdas, na grade de penalização efetivamente treinada





AMOSTRA REDUZIDA DO TREINO (600 CONTRATOS)

$$L_{pen} = L + \frac{1}{2C}\sum_j \beta_j^2$$

C e força	**0,0100 · 100**
log loss treino	**0,32275**
log loss validação	**0,35472**
AUC validação	**0,7170**

Penalização, apenas nos valores treinados

C = 0,0100

Com 600 contratos e a especificação flexível, a perda de treino continua caindo enquanto a de validação sobe. O mínimo de validação ocorre em C = 0,0100, com log loss 0,35472.

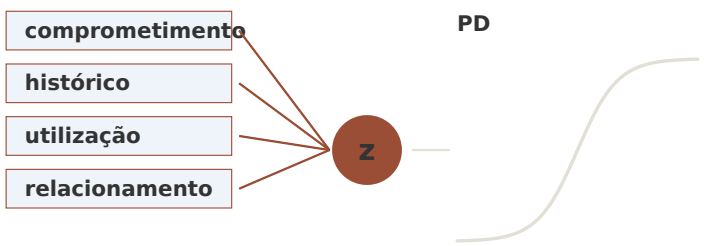
O teste final não participa desta escolha.

A melhor explicação do treino pode ser sensível demais para funcionar fora dele.

Experimento sintético; semente e versões no notebook

Explicar a estrutura é vantagem; especificar a estrutura é responsabilidade

Adequação do logit em três situações práticas



FORÇAS

- função compacta e auditável
- interpretação condicionada à codificação
- previsão rápida e fácil de implantar
- regularização e transformações disponíveis

LIMITES

- a forma funcional é escolhida, não descoberta
- variáveis ausentes ou mal medidas continuam ausentes
- extrapolação fora da faixa observada é frágil
- correlação entre atributos dificulta a leitura de um coeficiente
- mudança de população desloca o nível da previsão

SITUAÇÃO

As características somam efeitos de forma estável e as faixas de risco crescem de forma regular.

PERGUNTA DE DIAGNÓSTICO

As curvas de risco por faixa são compatíveis com uma soma de efeitos na escala do escore?

> **Ação possível.** Manter o logit como referência, documentar a especificação e verificar calibração fora do tempo.

O logit oferece uma estrutura explícita, mas a qualidade depende da especificação e da validação.

Síntese conceitual aplicada; sem dados

Bruno aumenta seu comprometimento de 38% para 48%. E agora?

Exercício: acompanhar a escala do escore e a escala da probabilidade

Bruno

perfil original

Renda	**R$ 4.500**
Comprometimento	**38%**
Relacionamento	**8 meses**
Utilização	**65%**
Histórico	**sim**
Canal	**Digital**

Uma única alteração: comprometimento de 38% para 48%.

REVISÃO RÁPIDA

A escolha feita aqui é preservada ao voltar.

ESCOLHA UMA ALTERNATIVA

A A PD sobe 10 pontos percentuais.

B O escore sobe 0,40 e as odds são multiplicadas por e elevado a 0,40.

C A PD é multiplicada por e elevado a 0,40.

Correta. A variação do escore é 0,04 × 10 = 0,40. As odds são multiplicadas por 1,4918 e a PD passa de 11,66% para 16,45%.

1. VARIAÇÃO DA CARACTERÍSTICA

comprometimento passa de 38% para 48%, ou seja Δcomp = +10 pontos

2. VARIAÇÃO DO ESCORE

Δz = 0,04 × 10 = 0,400, de −2,025 para −1,625

3. EFEITO NAS ODDS

odds passam de 0,1320 para 0,1969, multiplicadas por 1,4918

4. NOVA PROBABILIDADE

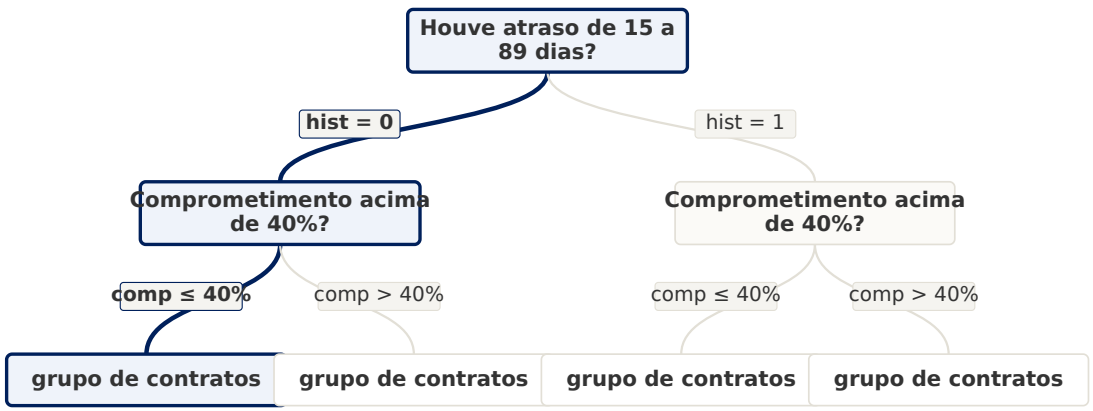
p = 1 / (1 + e^1,625) = 16,4516%, diferença de +4,79 p.p.

Entender o logit exige acompanhar a escala do escore e a escala da probabilidade.

Exemplo didático de logit; cálculo próprio

Houve atraso? Qual é o comprometimento?

Uma árvore organiza a previsão em perguntas sucessivas



As perguntas e seus limites são aprendidos dos dados. Aqui usamos uma árvore pequena para enxergar o mecanismo. As folhas ainda não mostram PD: primeiro a estrutura.

A árvore encaminha clientes semelhantes para uma mesma estimativa de risco.

Árvore didática de 1.000 contratos; cálculo próprio

Ana

Comprometimento	**22%**
Histórico	**não**

Só estas duas entram nas perguntas desta árvore.

Cliente

PERCURSO

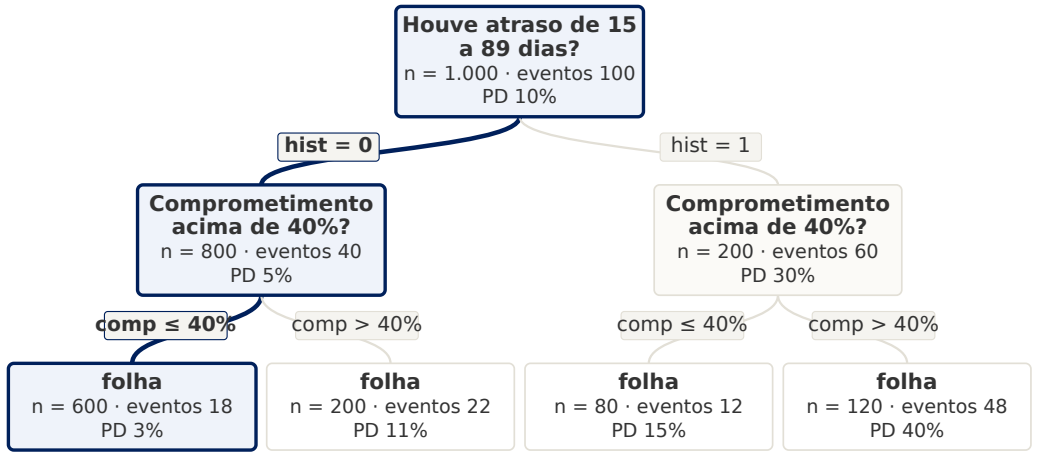
1. A raiz pergunta pelo histórico.
2. Resposta de Ana: hist = 0.
3. Segundo nó: comp ≤ 40%. Chega à folha.

> Regra completa da folha: hist = 0 e comp ≤ 40%.

Raiz: primeira pergunta. Nó: qualquer pergunta. Ramo: resposta que leva adiante. Folha: grupo final, onde a previsão é atribuída.

Nesta folha, 18 de 600 contratos ficaram inadimplentes

A PD da folha é a frequência observada no grupo



FOLHA	N	EVENTOS	PD
hist = 0 e comp ≤ 40%	**600**	**18**	**3%**
hist = 0 e comp > 40%	200	22	11%
hist = 1 e comp ≤ 40%	80	12	15%
hist = 1 e comp > 40%	120	48	40%
Soma das folhas	1.000	100	10%

A previsão é a frequência observada no grupo, sujeita à incerteza e à generalização.

Árvore didática de 1.000 contratos; cálculo próprio

FOLHA DE ANA

árvore didática

hist = 0 e comp ≤ 40%

$$\hat{p} = \frac{d}{n} = \frac{18}{600} = 3\%$$

3,0%

Barra vermelha: contratos com evento. Barra clara: contratos sem evento.

Cliente

CONFERÊNCIA

As quatro folhas somam 1.000 contratos e 100 eventos, exatamente os totais da raiz.

Mesma folha, mesma previsão: dois clientes com rendas diferentes recebem a mesma PD, porque a renda não participa das perguntas desta árvore.

Qual pergunta separa melhor os comportamentos observados?

Dois candidatos à primeira divisão, a partir da mesma raiz

Raiz 1.000 contratos · 100 eventos · taxa 10% **10,0%**

Histórico de atraso

hist = 0 n = 800 · eventos 40 · taxa 5,00%

hist = 1
n = 200 · eventos 60 · taxa 30,00%

Comprometimento até 40%

comp ≤ 40% n = 680 · eventos 30 · taxa 4,41%

comp > 40%
n = 320 · eventos 70 · taxa 21,88%

EXPLORAÇÃO

Regra de escolha: a divisão escolhida é a de menor impureza ponderada, considerando ao mesmo tempo a mistura de classes em cada lado e o tamanho de cada lado.

O cálculo completo com o índice de Gini vem no próximo slide.

Os dois candidatos partem exatamente da mesma raiz: 1.000 contratos e 100 eventos.

Uma boa divisão cria grupos mais homogêneos segundo a função objetivo.

Árvore didática de 1.000 contratos; cálculo próprio

Histórico reduz o Gini de 0,180 para 0,160

Impureza da raiz, média ponderada dos filhos e ganho

1. IMPUREZA DA RAIZ

$$p = \frac{100}{1.000} = 10\%$$
$$G(\text{raiz}) = 2\,p\,(1-p) = 0{,}180$$

2. IMPUREZA DE CADA FILHO

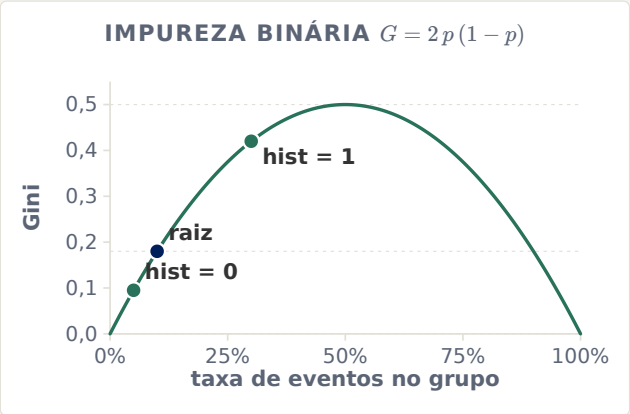
$$G_{\text{esq}} = 2\,p_{\text{esq}}(1-p_{\text{esq}}) = 0{,}0950 \quad \text{hist} = 0,\ p_{\text{esq}} = 5{,}00\%$$
$$G_{\text{dir}} = 2\,p_{\text{dir}}(1-p_{\text{dir}}) = 0{,}4200 \quad \text{hist} = 1,\ p_{\text{dir}} = 30{,}00\%$$

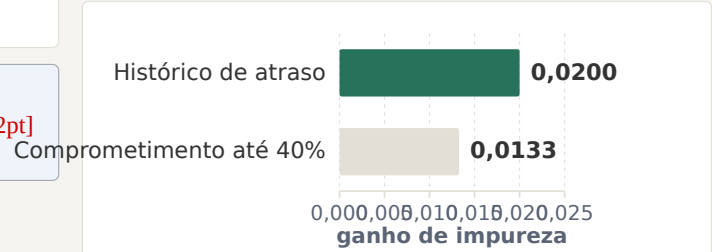
3. PONDERAÇÃO PELO TAMANHO

$$\begin{aligned}\bar{G} &= \frac{n_{\text{esq}}}{n}\,G_{\text{esq}} + \frac{n_{\text{dir}}}{n}\,G_{\text{dir}} \\ &= \frac{800}{1.000}\cdot 0{,}0950 + \frac{200}{1.000}\cdot 0{,}4200 \\ &= 0{,}160000\end{aligned}$$

4. GANHO E COMPARAÇÃO

\begin{aligned}\Delta G &= G(\text{raiz}) - \bar{G} = 0{,}180 - 0{,}160000 = 0{,}020000 \\[2pt] &\text{candidato comprometimento até 40%: } \Delta G = 0{,}013272\end{aligned}



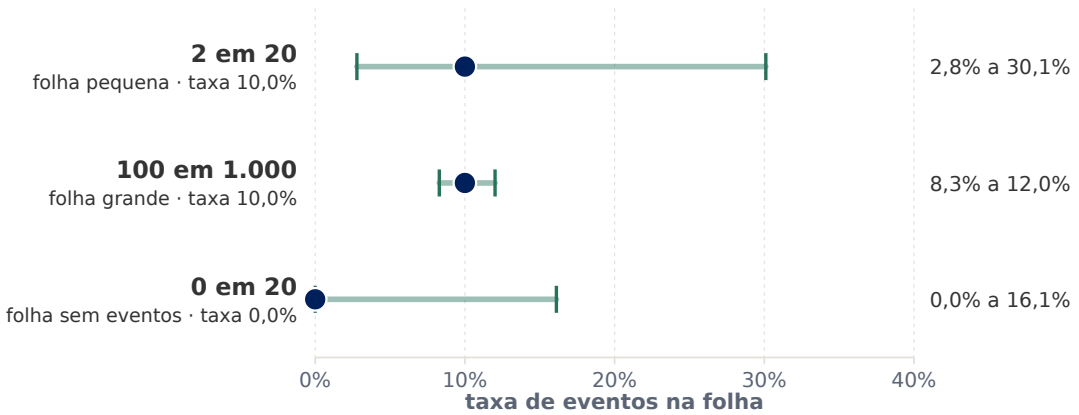


Não basta uma folha pura: importa como toda a amostra fica depois da divisão.

Árvore didática de 1.000 contratos; cálculo próprio

Duas folhas com PD de 10%. A evidência é igual?

Intervalo binomial descritivo de Wilson, nominal de 95%



FOLHA	EVENTOS	N	TAXA	INTERVALO DE WILSON
folha pequena	2	20	10,0%	2,8% a 30,1%
folha grande	100	1.000	10,0%	8,3% a 12,0%
folha sem eventos	0	20	0,0%	0,0% a 16,1%

10% em 20 contratos não oferece a mesma precisão que 10% em 1.000.

Pares binomiais ilustrativos; cálculo próprio

LEITURA

$\text{amplitude} \propto \sqrt{\hat{p}(1-\hat{p})/n}$ Mesma taxa, evidências diferentes: com 20 contratos a amplitude é 27,3 p.p.; com 1.000, 3,7 p.p..

Tamanho da folha pequena, mantendo a taxa em 10%

2 em 20

RESSALVA

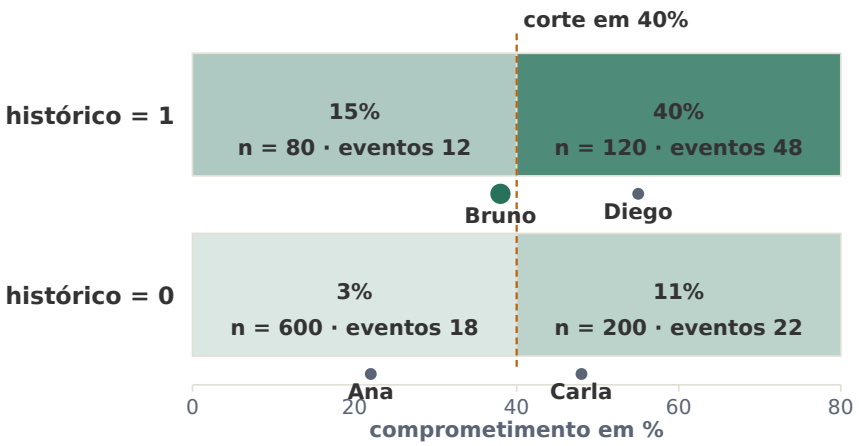
Intervalo binomial descritivo. Não incorpora a escolha adaptativa das folhas.

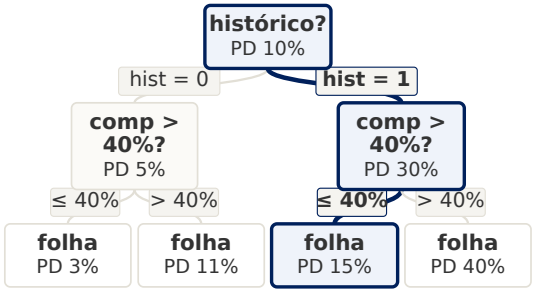
Uma folha com zero eventos não prova risco zero: o limite superior permanece positivo.

É por isso que a árvore recebe um tamanho mínimo de folha como parâmetro.

O mesmo aumento de comprometimento pode mudar de folha

As regras da árvore dividem o espaço em regiões de previsão constante





Cliente **perfil original**

Comprometimento

PD 15,0%

Regra atual: hist = 1 e comp ≤ 40%, PD 15%.

Em comprometimento exatamente igual a 40, o cliente segue pelo ramo até 40.

Os efeitos dependem do caminho. Dentro de uma folha, a previsão é constante.

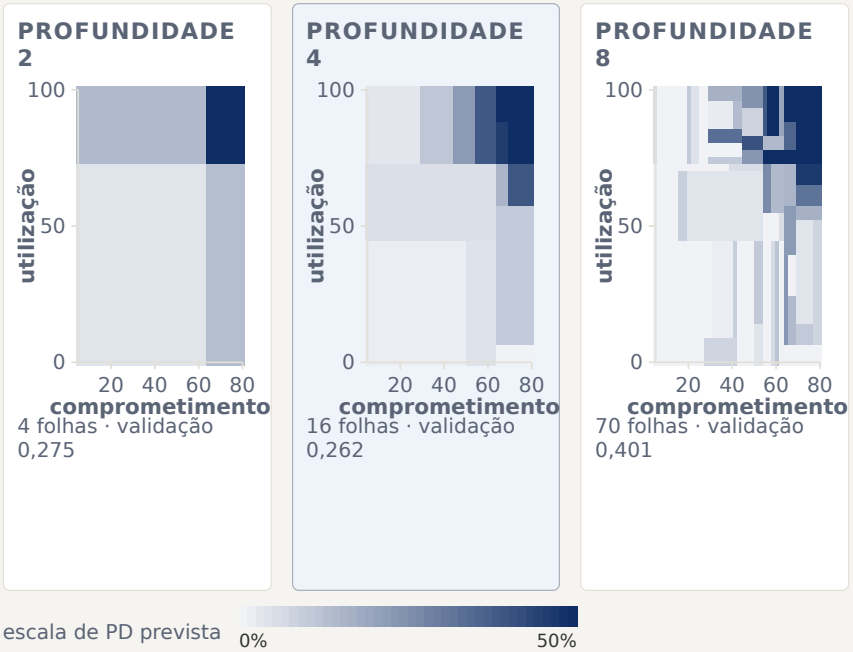
Árvore didática de 1.000 contratos; cálculo próprio

A árvore cresce. O resultado fora do treino acompanha?

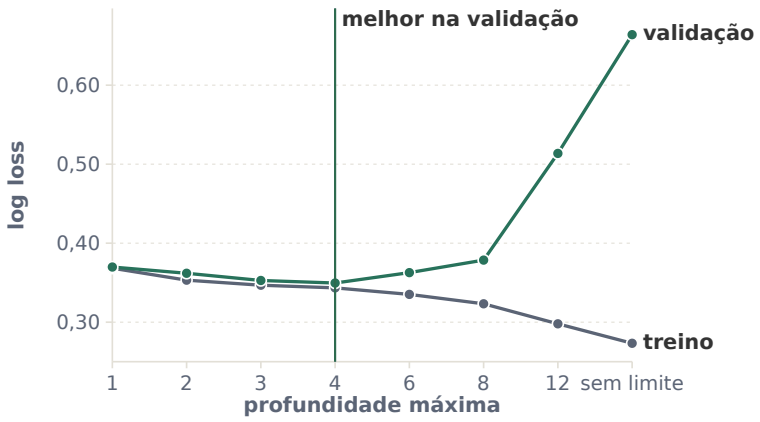
Profundidade, perda de treino e perda de validação no experimento sintético

EXPERIMENTO AUXILIAR DE DUAS VARIÁVEIS: AS REGIÕES APRENDIDAS

experimento auxiliar, não o modelo completo



BASE PRINCIPAL: PERDA POR PROFUNDIDADE, COM MÍNIMO DE 20 CONTRATOS POR FOLHA



Profundidade máxima, apenas nos ajustes executados

4

16 folhas mediana de 366 contratos por folha

treino 0,3434 · validação 0,3496

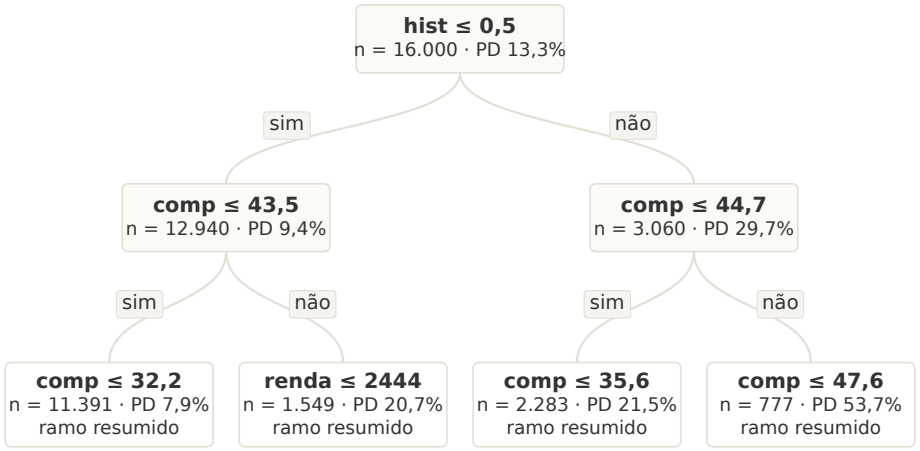
Uma árvore que memoriza o treino pode produzir probabilidades frágeis fora dele.

Experimento sintético; semente e versões no notebook

Três controles, três formas de conter o excesso de divisões

Profundidade máxima, mínimo de contratos por folha e poda por complexidade

MODELO SELECIONADO, RESUMIDO ATÉ O NÍVEL 2

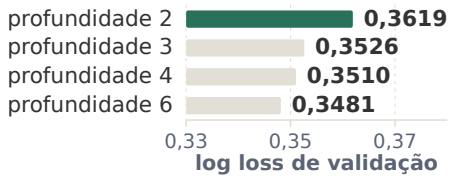


Limita quantas perguntas em sequência a árvore pode fazer. Aqui o mínimo por folha fica fixo em 300 contratos.

Cenário treinado
profundidade 2

4 folhas mediana de 1.916 contratos por folha

validação 0,3619



Nomes abreviados nas caixas: comp é comprometimento, hist é histórico, util é utilização, rel é relacionamento. Ramos além do nível mostrado aparecem resumidos. A árvore selecionada tem 24 folhas.

Controlar complexidade é escolher quais detalhes dos dados vale a pena manter.

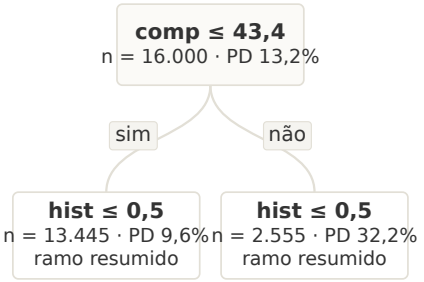
Experimento sintético; semente e versões no notebook

Se mudarmos um pouco a amostra, a árvore continua contando a mesma história?

Vinte reamostragens do treino, com os mesmos hiperparâmetros

RÉPLICA A: Nº 1

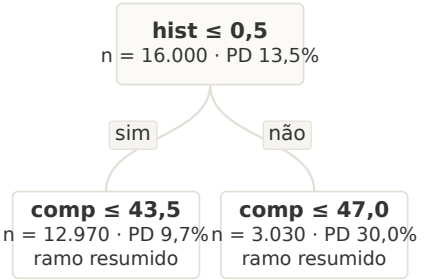
8 folhas · AUC validação 0,7028



nomes abreviados nas caixas. raiz: comprometimento ≤ 43,35 · PD de Bruno: 14,85%

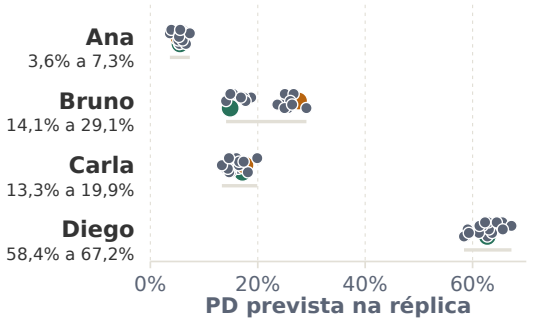
RÉPLICA B: Nº 3

7 folhas · AUC validação 0,7065



nomes abreviados nas caixas. raiz: histórico ≤ 0,50 · PD de Bruno: 27,59%

PD NAS 20 RÉPLICAS



Réplica A

Réplica B

Mesma complexidade nas 20 réplicas: profundidade 3 e mínimo de 300 contratos por folha.

A raiz mudou de comprometimento para histórico apenas por reamostragem do treino.

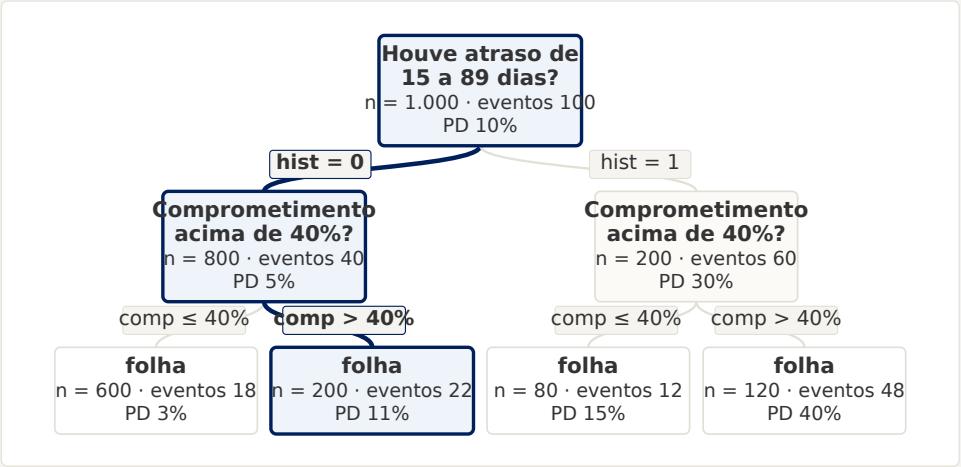
Entre as 20 réplicas, a raiz foi comprometimento em 10 delas e histórico em 10 delas.

Uma regra fácil de ler pode ser sensível à amostra que a produziu.

Experimento sintético; semente e versões no notebook

Carla recebe qual PD nesta árvore?

Exercício: percorrer, calcular e questionar



Carla

R$ 10.000 por mês · comprometimento 48% · relacionamento 60 meses · utilização 85% · histórico não · canal digital

1. INDIQUE O CAMINHO

Caminho correto: Carla tem histórico 0 e comprometimento 48%, acima de 40%.

2. CALCULE A PD DA FOLHA

em porcentagem

Correto: 22 dividido por 200 resulta em 11%.

3. SE A RENDA DE CARLA DOBRASSE, MANTENDO FIXAS AS ENTRADAS USADAS NA ÁRVORE

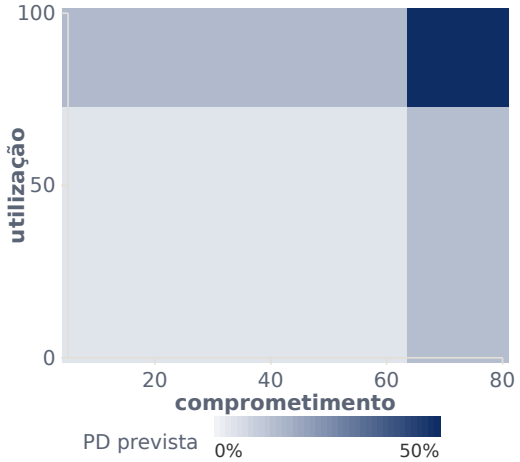
Uma regra legível ajuda a explicar o cálculo, mas não dispensa evidência de qualidade.

Árvore didática de 1.000 contratos; cálculo próprio

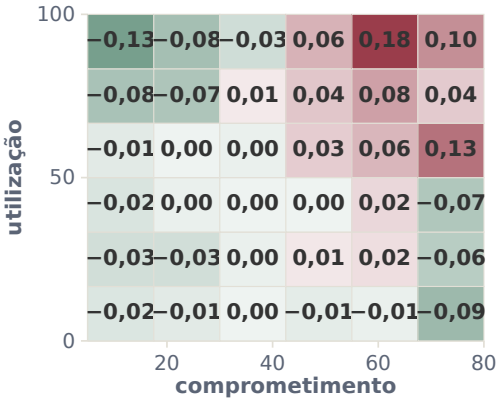
E se uma árvore pequena ainda deixar estrutura nos erros?

Resíduos agregados por região, no experimento auxiliar de duas variáveis

PREVISÃO DE UMA ÁRVORE DE PROFUNDIDADE 2



MÉDIA DE Y − P POR REGIÃO



Vermelho: a previsão ficou baixa demais. Verde: ficou alta demais.

O QUE SOBRA DEPOIS DA ÁRVORE RASA

regiões com resíduo positivo	**13 de 36**
média dessas regiões	**0,0567**
média das demais	**−0,0325**
validação da árvore rasa	**0,2752**

Os resíduos mostrados são do conjunto usado no ajuste desta árvore auxiliar, com duas variáveis. Nenhum resultado do teste participa.

Em vez de exigir tudo de uma árvore, podemos construir a previsão por acréscimos.

Experimento auxiliar de duas variáveis; cálculo próprio

A próxima árvore aprende a partir das previsões atuais

Um ciclo de atualização em quatro passos

novo escore = escore atual + taxa de aprendizagem × contribuição da árvore

$$F_m = F_{m-1} + \eta\, h_m$$

1. PREVER COM O QUE JÁ EXISTE

$$p = \sigma(F) = \frac{1}{1+e^{-F}} = \frac{1}{1+e^{1{,}3863}} = 20{,}0000\%$$

A previsão atual do grupo vem do escore acumulado até aqui, na mesma logística do slide 09.

2. CALCULAR A DIREÇÃO DE MELHORIA

$$r = -\frac{\partial L}{\partial F} = y - p = 0 - 0{,}2000 = -0{,}2000$$

O resíduo não é escolha de conveniência: para a perda logística do slide 13, o gradiente negativo em relação ao escore é exatamente y menos p. Por isso a técnica se chama gradient boosting.

3. AJUSTAR UMA ÁRVORE PEQUENA AOS R

$$h_1(\text{folha B}) = \bar{r}_{\text{B}} = 0{,}200000$$

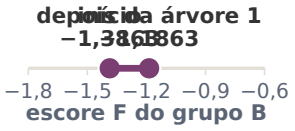
A árvore devolve um número por folha, positivo ou negativo. Não é um voto.

4. ATUALIZAR O ESCORE

$$F_1 = F_0 + \eta\, h_1 = -1{,}3863 + 1 \times 0{,}200000 = -1{,}1863$$

Só depois de somar todas as contribuições aplicamos a função logística.

A RÉGUA DO ESCORE ACUMULADO



PD aparece apenas depois da transformação logística: 20,00% passa a 23,39%

Miniatura didática de 10 registros, com taxa de aprendizagem igual a 1 apenas para facilitar a leitura. O registro usado no passo 2 tem y = 0 e pertence ao grupo B.

Cada árvore responde ao que ainda precisa melhorar na previsão atual.

Miniatura didática de 10 registros; cálculo próprio

Dois eventos em dez contratos: começamos em 20%

O melhor preditor constante da perda logarítmica no treino

previsão inicial 20,00% $F_0 = \ln\left(\frac{0{,}20}{0{,}80}\right) = -1{,}3863$ perda média 0,5004

miniatura didática, distinta da base de comparação

GRUPO A

comprometimento 30% · 0 eventos em 5

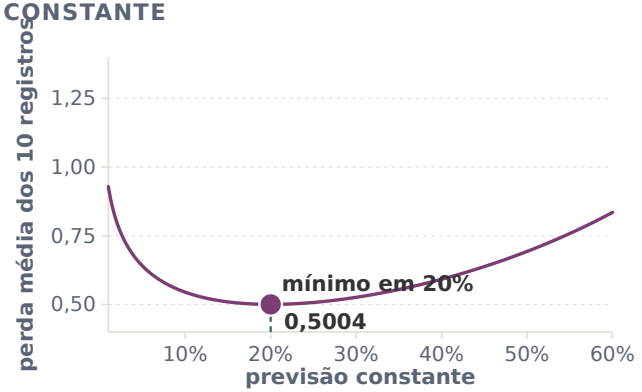
REGISTRO	y	p_0	$r = y - p_0$
nº 1	0	20,00%	
nº 2	0	20,00%	
nº 3	0	20,00%	
nº 4	0	20,00%	
nº 5	0	20,00%	

GRUPO B

comprometimento 50% · 2 eventos em 5

REGISTRO	y	p_0	$r = y - p_0$
nº 6	0	20,00%	
nº 7	0	20,00%	
nº 8	0	20,00%	
nº 9	1	20,00%	
nº 10	1	20,00%	

PERDA MÉDIA EM FUNÇÃO DA PREVISÃO CONSTANTE



Previsão constante

perda 0,5004

A curva é calculada diretamente com os dez rótulos, sem treinar biblioteca alguma. O mínimo ocorre na média empírica de y, igual a 20%.

Antes de usar características, a taxa observada fornece o ponto de partida do exemplo.

Miniatura didática de 10 registros; cálculo próprio

A primeira correção reduz A e aumenta B

Resíduos, média por folha e atualização do escore

OS DEZ REGISTROS

primeira ordem didática

REGISTRO	GRUPO	y	p_{atual}	$r = y - p$	$h(\text{folha})$	F_{novo}	p_{novo}
nº 1	A	0	20,00%	−0,2000	−0,200000	−1,586294	16,9906%
nº 2	A	0	20,00%	−0,2000	−0,200000	−1,586294	16,9906%
nº 3	A	0	20,00%	−0,2000	−0,200000	−1,586294	16,9906%
nº 4	A	0	20,00%	−0,2000	−0,200000	−1,586294	16,9906%
nº 5	A	0	20,00%	−0,2000	−0,200000	−1,586294	16,9906%
nº 6	B	0	20,00%	−0,2000	0,200000	−1,186294	23,3922%
nº 7	B	0	20,00%	−0,2000	0,200000	−1,186294	23,3922%
nº 8	B	0	20,00%	−0,2000	0,200000	−1,186294	23,3922%
nº 9	B	1	20,00%	0,8000	0,200000	−1,186294	23,3922%
nº 10	B	1	20,00%	0,8000	0,200000	−1,186294	23,3922%

destacar um registro:

A ÁRVORE AJUSTADA AOS RESÍDUOS

POR GRUPO	A	B
escore inicial	−1,386294	−1,386294
contribuição da árvore	−0,200000	+0,200000
escore novo	−1,586294	−1,186294
PD	16,9906%	23,3922%

etapa 4 de 4

A árvore aprende um padrão nos gradientes e aplica a mesma correção a quem cai na mesma folha.

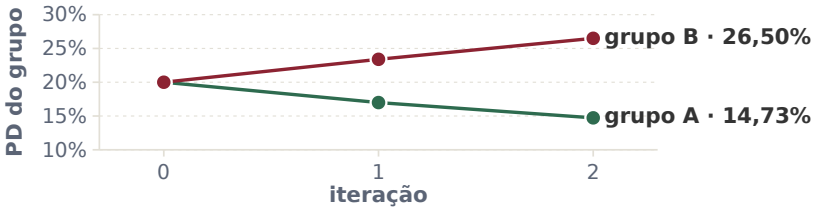
Miniatura didática de 10 registros; cálculo próprio

A primeira árvore mudou as previsões. Agora mudam os gradientes

A segunda correção é recalculada depois da primeira

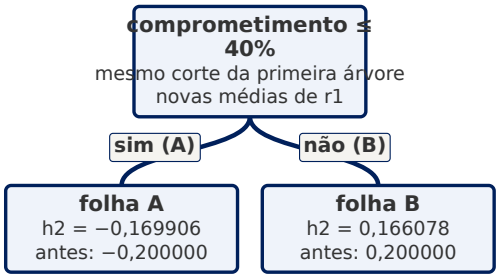
OS MESMOS DEZ REGISTROS

REGISTRO	GRUPO	Y	P0	R0	P1	R1	F2	P2
nº 1	A	0	20,00%	−0,2000	16,9906%	−0,169906	−1,756200	14,7267%
nº 2	A	0	20,00%	−0,2000	16,9906%	−0,169906	−1,756200	14,7267%
nº 3	A	0	20,00%	−0,2000	16,9906%	−0,169906	−1,756200	14,7267%
nº 4	A	0	20,00%	−0,2000	16,9906%	−0,169906	−1,756200	14,7267%
nº 5	A	0	20,00%	−0,2000	16,9906%	−0,169906	−1,756200	14,7267%
nº 6	B	0	20,00%	−0,2000	23,3922%	−0,233922	−1,020217	26,4985%
nº 7	B	0	20,00%	−0,2000	23,3922%	−0,233922	−1,020217	26,4985%
nº 8	B	0	20,00%	−0,2000	23,3922%	−0,233922	−1,020217	26,4985%
nº 9	B	1	20,00%	0,8000	23,3922%	0,766078	−1,020217	26,4985%
nº 10	B	1	20,00%	0,8000	23,3922%	0,766078	−1,020217	26,4985%



No grupo B, o resíduo de quem teve evento caiu de 0,8 para 0,766078 e o de quem não teve ficou mais negativo, de −0,2 para −0,233922.

A SEGUNDA ÁRVORE



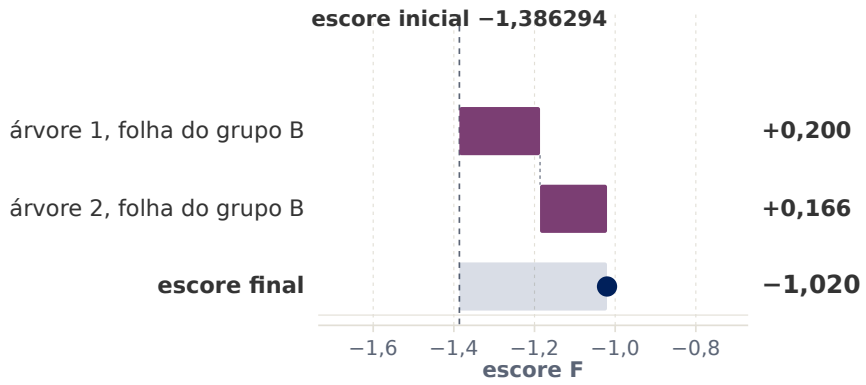
A exposição principal para em duas iterações. A correção média do grupo B continua positiva, mas caiu de 0,200000 para 0,166078.

A próxima correção é recalculada depois da anterior. Ela não repete mecanicamente o mesmo passo.

Miniatura didática de 10 registros; cálculo próprio

Escore inicial mais árvore 1 mais árvore 2 é a nova PD

Somamos contribuições no escore e só então calculamos a probabilidade



$$F = -1{,}020217 \quad p = \sigma(F) = 26{,}4985\%$$

grupo A chegaria a −1,756200 e PD 14,7267%

SOLICITANTE NOVO

sem desfecho conhecido

Comprometimento do solicitante

grupo B

Nenhum campo de desfecho aparece nesta ficha. O y foi usado apenas no treinamento.

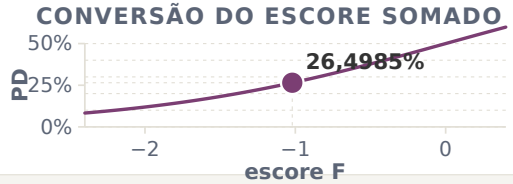
árvore 1 **+0,200000**

comprometimento 50% é maior que 40, portanto folha do grupo B

árvore 2 **+0,166078**

comprometimento 50% é maior que 40, portanto folha do grupo B

CONVERSÃO DO ESCORE SOMADO

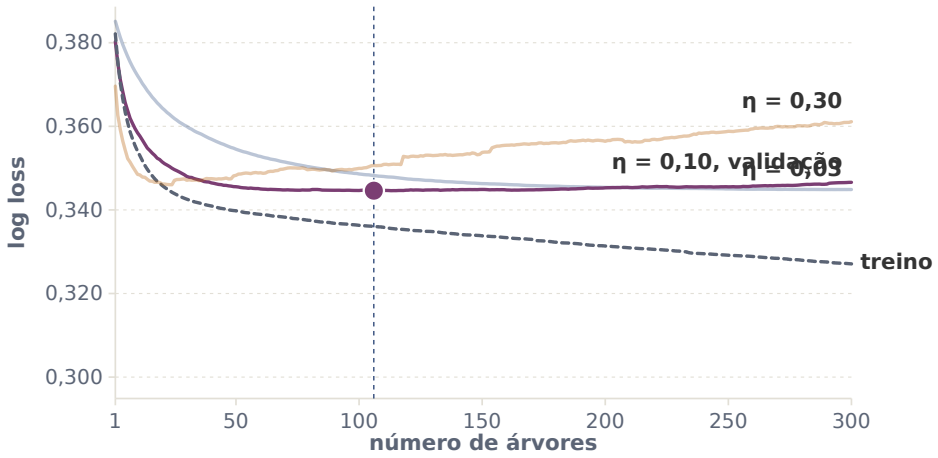


Para prever, o cliente percorre as árvores já aprendidas. Não precisamos conhecer seu desfecho.

Miniatura didática de 10 registros; cálculo próprio

Learning rate e número de árvores precisam ser escolhidos juntos

Trajetórias de perda treinadas com cada taxa, no experimento sintético



$$F_m = F_{m-1} + \eta\, h_m$$

taxa de aprendizagem	**0,10**
árvores usadas	**106**
log loss de validação	**0,34460**
melhor iteração desta taxa	**106 com 0,34460**

Taxa de aprendizagem treinada

Número de árvores, apenas em iterações calculadas

106 árvores · 0,34460

PD com 105 árvores: **Ana 5,00%** **Bruno 23,35%** **Carla 17,60%** **Diego 61,05%**

Trajetórias calculadas a cada cinco iterações e registradas no notebook.

Mudar o passo altera previsões, gradientes e, portanto, o que as próximas árvores aprendem.

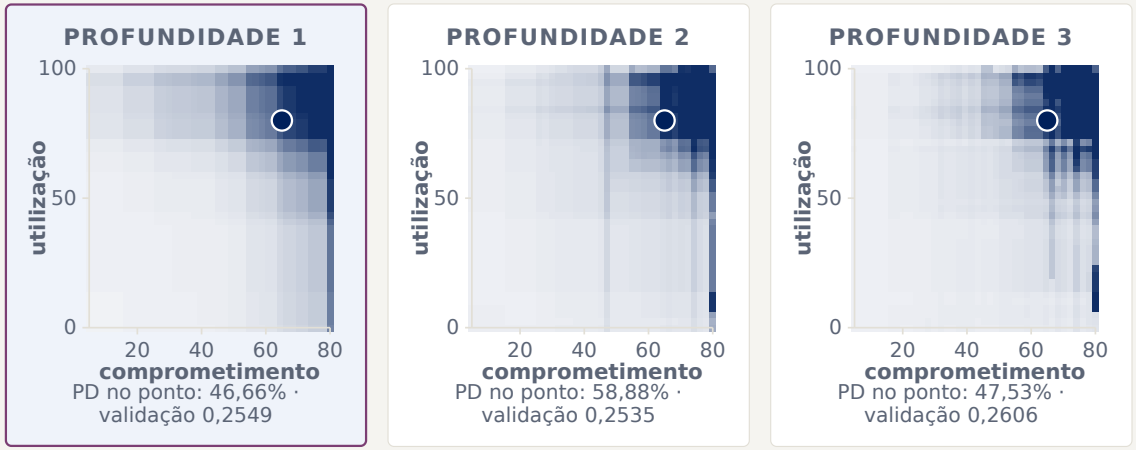
Experimento sintético; semente e versões no notebook

Uma correção pode depender de uma característica ou da combinação de várias

Profundidade de cada árvore no experimento auxiliar de duas variáveis

MESMA ESCALA E MESMO DOMÍNIO NOS TRÊS PAINÉIS

experimento auxiliar de duas variáveis



escala de PD prevista 0% 60%

A regra do canto alto e alto: se comprometimento acima de 40% e utilização acima de 50%, o risco sobe mais do que a soma dos dois efeitos isolados. Taxa 0,1 e 150 árvores nos três painéis; muda apenas a profundidade.

Cliente sintético: comprometimento 65%, utilização 80%, PD 46,66% na profundidade 1.

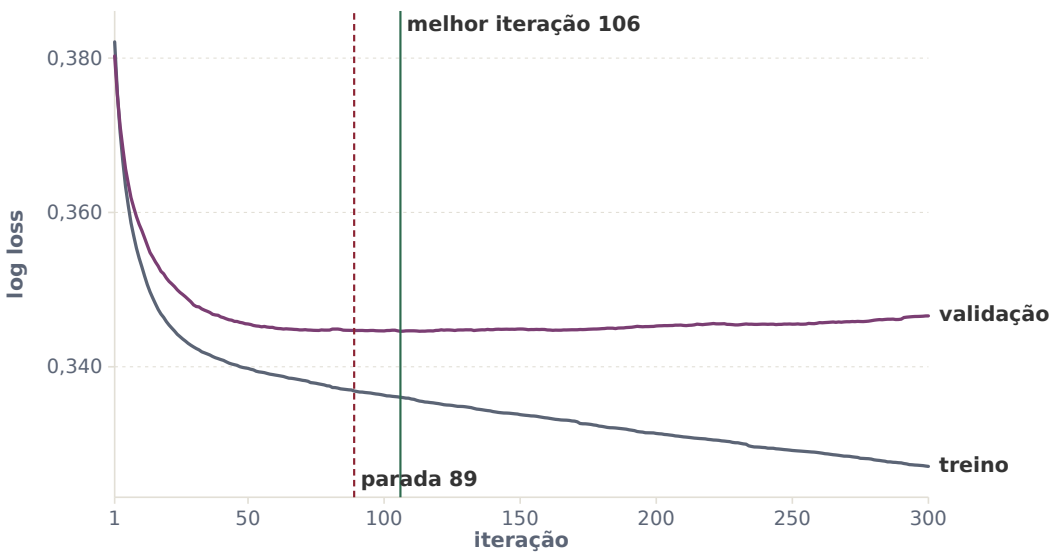
Comprometimento
65%
Utilização
80%

Muitas regras simples somadas não são sempre equivalentes a regras que combinam condições.

Experimento auxiliar de duas variáveis; cálculo próprio

Continuar melhorando o treino pode deixar de ajudar a validação

Regra de parada com tolerância e paciência, definida antes de olhar o teste



REGRA DEFINIDA ANTES DE TREINAR

máximo de 300 iterações · tolerância 0,000100 · paciência 20 · escolha na validação fora do tempo

300 de 300 iterações reveladas · melhor validação até aqui 0,34467 · paciência 203 de 20

Curvas calculadas no notebook, reveladas em lotes de 20 iterações.

MELHOR ITERAÇÃO E ITERAÇÃO DE PARADA

Melhor iteração na validação: 106, com log loss 0,34460. A paciência encerraria em 89, depois do mínimo. O modelo servido é o da melhor iteração.

A curva do teste não aparece em nenhum seletor.

A melhor iteração é escolhida usando uma amostra destinada a essa escolha.

Experimento sintético; semente e versões no notebook

Mais flexibilidade exige mais disciplina de validação

Quatro situações, a pergunta de diagnóstico e a evidência que a responde

SINAL APROXIMADAMENTE ADITIVO

As características contribuem de forma estável e quase somável na escala do escore.

cuidado: Complexidade sem ganho é custo de manutenção sem retorno.

INTERAÇÕES RELEVANTES

O risco depende de combinações de condições, como comprometimento alto junto de utilização alta.

cuidado: Encontrar interação no treino é fácil. Reproduzir a interação fora do tempo é o teste.

POUCOS EVENTOS

Volume pequeno de inadimplências, faixas finas e estimativas instáveis.

cuidado: Flexibilidade com poucos eventos costuma ajustar ruído.

MUDANÇA DE POPULAÇÃO

A composição dos solicitantes muda ao longo do tempo.

cuidado: Nenhum algoritmo corrige sozinho mudança de população.

GANHOS POSSÍVEIS

- relações não lineares sem especificação manual
- interações capturadas pela profundidade das árvores
- boa adaptação a dados tabulares com muitas colunas

CUSTOS

- ajuste de hiperparâmetros e orçamento de busca
- explicação menos direta que um coeficiente
- risco de sobreajuste quando a validação é frouxa
- manutenção, versionamento e verificação de calibração

O QUE VERIFICAR EM SINAL APROXIMADAMENTE ADITIVO

O boosting ganha da especificação linear bem construída, fora do tempo?

> **Qual evidência aceitar.** Aceitar apenas ganho que persiste na validação fora do tempo contra um logit com transformações e interações plausíveis. Se o ganho não persiste, o logit basta.

SENSIBILIDADE JÁ CALCULADA

validação, não o teste

MODELO	AUC VALIDAÇÃO	LOG LOSS VALIDAÇÃO
Logit regularizado	0,7379	0,34299
Logit flexível	0,7385	0,34395
Gradient boosting	0,7366	0,34460

Neste experimento os três ficam próximos na validação. Isso é o resultado observado aqui, não uma generalização de mercado.

Capturar padrões complexos é uma capacidade. Demonstrar que eles persistem é uma obrigação.

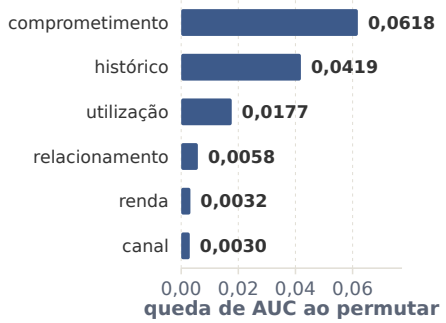
Síntese conceitual, com um resultado calculado como exemplo

O que o modelo usa em geral? O que pesou para Bruno?

Importância global por permutação e contribuição local no escore

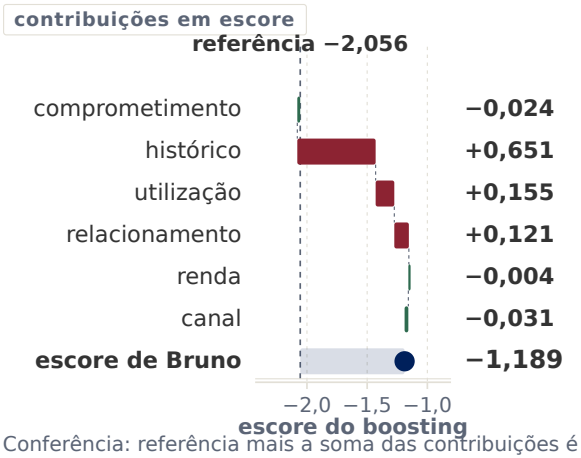
GLOBAL: IMPORTÂNCIA NA CARTEIRA

validação, jan a jun de 2021



AUC de referência na validação: 0,7366. A barra mostra quanto a AUC cai quando os valores da variável são embaralhados.

LOCAL: O ESCORE DE BRUNO



Conferência: referência mais a soma das contribuições é −1,189179, e o escore do modelo é −1,189179. Diferença menor que 10 elevado a menos 9.

DO ESCORE À PD

escore do boosting	**−1,1892**
PD do modelo	**23,34%**
PD depois da calibração	**25,78%**

Cliente

Destacar uma variável (nome abreviado)

COMO LER

O painel global mede desempenho: quanto a AUC cai ao embaralhar a variável. O local mede escore: quanto a característica desloca a previsão daquele cliente. Alterar a variável não garante essa redução: explicação não identifica causa.

Uma variável importante na carteira não precisa ser o principal fator da previsão de todo cliente.

Experimento sintético; semente e versões no notebook

O novo cliente caiu no grupo B. Qual é a previsão?

Exercício: somar em escore e converter uma única vez

O QUE JÁ SABEMOS

escore inicial F0	**−1,386294**
folha da árvore 1	**+0,200000**
folha da árvore 2	**+0,166078**
taxa de aprendizagem	**1**
PD final	**26,4985%**

PERGUNTA DE INTERPRETAÇÃO

Uma contribuição positiva do histórico prova que remover o atraso do cadastro reduziria o risco real?

> Não. A contribuição explica a função preditiva aprendida com registros históricos. Ela não identifica o efeito de uma intervenção sobre o cadastro.

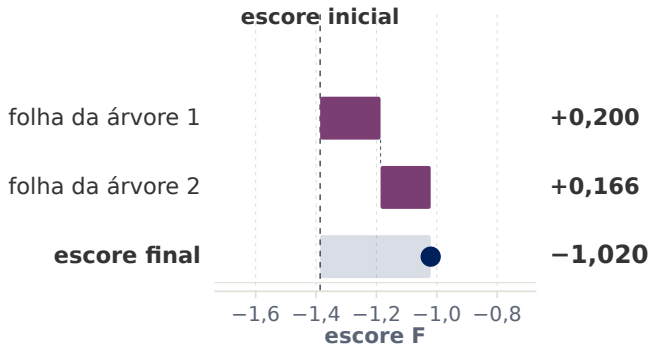
ESCOLHA O MÉTODO CORRETO

A Somar como pontos percentuais: 20% mais 20% mais 16,6078%.

B Somar ao escore e aplicar a função logística uma única vez.

C Tirar a média das duas folhas.

> **Correta.** Escore final −1,386294 mais 0,200000 mais 0,166078 é igual a −1,020217. Aplicando a logística, PD de 26,4985%.



26,4985%

sigmoid(−1,020217)

Com as árvores treinadas, os atributos determinam o percurso e a soma determina a PD.

Miniatura didática de 10 registros; cálculo próprio

Estamos comparando modelos ou comparando condições diferentes?

O protocolo precede qualquer métrica

CONDIÇÃO	LOGIT REGULARIZADO	ÁRVORE CONTROLADA	GRADIENT BOOSTING
Amostra e alvo mesmas coortes, mesmo evento e mesmo horizonte de 12 meses	**comum**		
Informação disponível as seis características, com a mesma regra de disponibilidade	**comum**		
Pré processamento tratamento próprio é permitido, sem vantagem de informação	padronização, mais a imputação comum	imputação pela mediana do treino e indicadores de ausência	
Escolha de complexidade sempre pela perda na validação fora do tempo	grade de penalização L2	profundidade, mínimo por folha e poda	taxa, profundidade e número de árvores
Avaliação final uma única passagem, com regras definidas antes	**teste congelado**		

OS QUATRO PERÍODOS, COM A MATURAÇÃO DE 12 MESES

treino 16.000 · validação 4.000 · calibração e política 3.000 · teste 5.000 contratos

DETECTAR COMPARAÇÃO INJUSTA

PARTIÇÃO	CONTRATAÇÕES	N	EVENTOS
treino	jan/2018 a dez/2019	16.000	2.124
validação	jan/2021 a jun/2021	4.000	522
calibração e política	ago/2022 a dez/2022	3.000	464
teste	fev/2024 a jul/2024	5.000	712

semente 20.260.920 · numpy 2.4.6 · sklearn 1.9.1 · python 3.11.15

sha256 do arquivo de resultados: 819746e8a01e72c2e9974080...

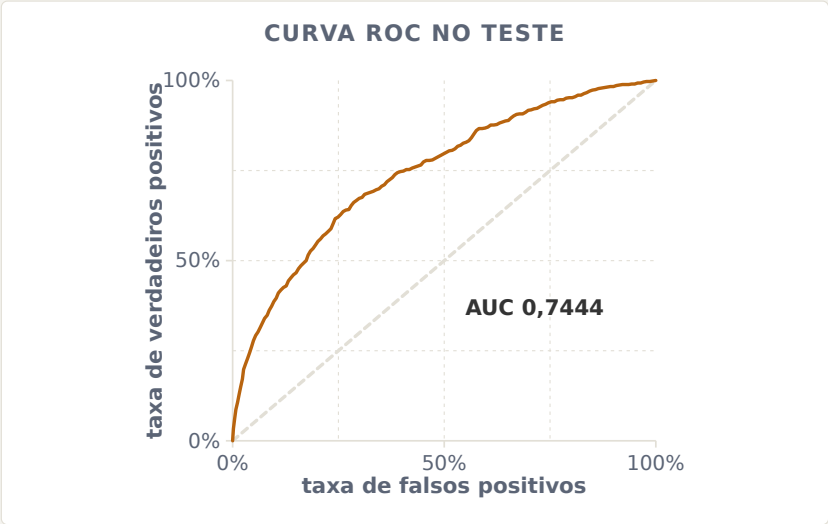
Mesmo alvo, mesma informação e teste preservado são condições da comparação.

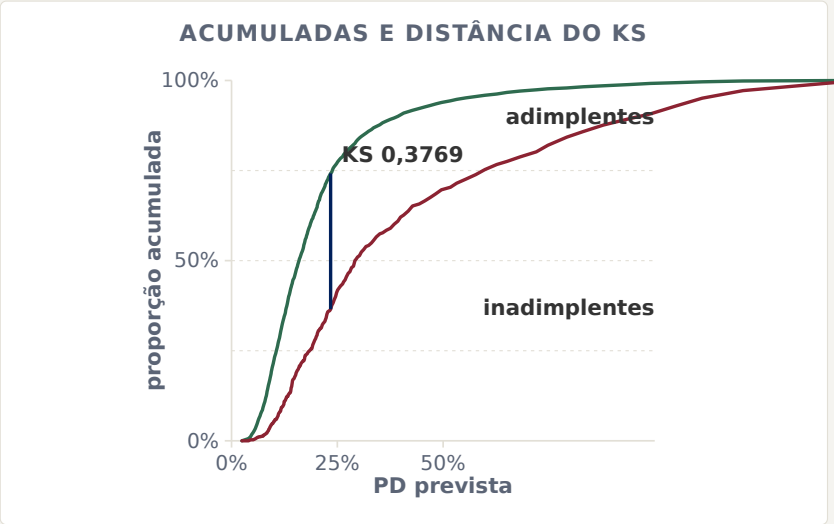
Experimento sintético; semente e versões no notebook

Quem aparece primeiro na fila de risco?

AUC contada por pares e curvas do experimento sintético

modelo teste, 5.000 contratos de fev a jul de 2024





AUC 0,7444 **KS 0,3769** n = 5.000 contratos, 712 eventos

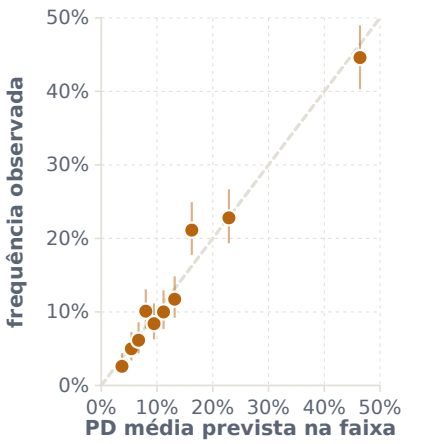
Mudar o corte não altera a AUC. A taxa de falsos positivos do eixo não é a inadimplência da carteira aprovada.

Um modelo pode ordenar bem mesmo que suas probabilidades estejam na escala errada.

Microexemplo de ranking e experimento sintético, em abas identificadas

Entre os clientes com PD próxima de 10%, quantos ficaram inadimplentes?

Média prevista contra frequência observada, faixa a faixa



FAIXA	N	EVENTOS	PREVISTA	OBSERVADA	WILSON
1	500	13	3,71%	2,60%	1,5% a 4,4%
2	501	25	5,40%	4,99%	3,4% a 7,3%
3	504	31	6,70%	6,15%	4,4% a 8,6%
4	495	50	7,98%	10,10%	7,7% a 13,1%
5	500	42	9,46%	8,40%	6,3% a 11,2%
6	500	50	11,17%	10,00%	7,7% a 12,9%
7	503	59	13,16%	11,73%	9,2% a 14,8%
8	497	105	16,24%	21,13%	17,8% a 24,9%
9	500	114	22,89%	22,80%	19,3% a 26,7%
10	500	223	46,40%	44,60%	40,3% a 49,0%

PREVISÕES ORIGINAIS

AUC	**0,744367**
Brier	**0,10676**
log loss	**0,35694**
PD média prevista	**14,31%**
taxa observada	**14,24%**

Modelo

Versão das previsões

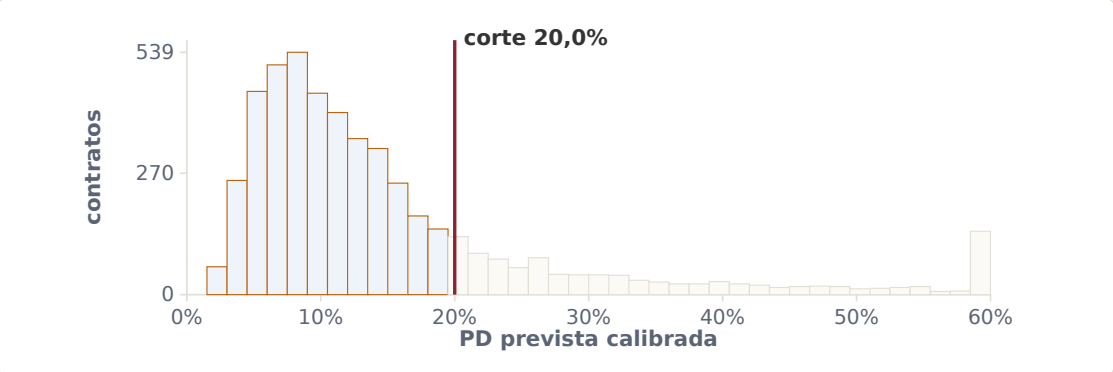
$\text{odds} \times 2 \iff z + \ln 2$: deslocamento constante no log odds não troca ninguém de lugar na fila, então a AUC permanece 0,744367. Brier e log loss mudam.

Calibrar é verificar a escala das probabilidades, não apenas a ordem dos clientes.

Experimento sintético; semente e versões no notebook

Até qual PD vamos aprovar?

O corte define quem entra na carteira e o que se observa depois



NO CORTE DE 20,0%

partição de política

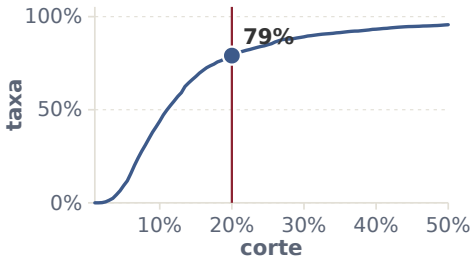
aprovados	**2.372 de 3.000**
taxa de aprovação	**79,1%**
eventos entre aprovados	**236**
inadimplência entre aprovados	**9,95%**
PD média dos aprovados	**9,86%**

Corte de aprovação

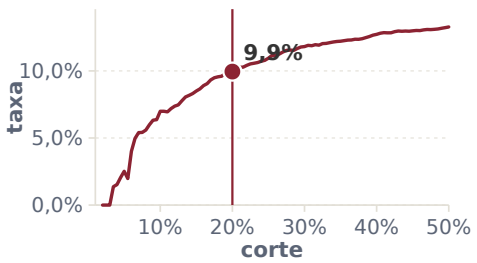
aprovação 79,1%

A política é definida nesta partição, antes de qualquer olhar ao teste.

TAXA DE APROVAÇÃO $\frac{\#\{\hat{p}\le c\}}{n}$



INADIMPLÊNCIA ENTRE APROVADOS $\frac{\sum_{\hat{p}\le c} y}{\#\{\hat{p}\le c\}}$

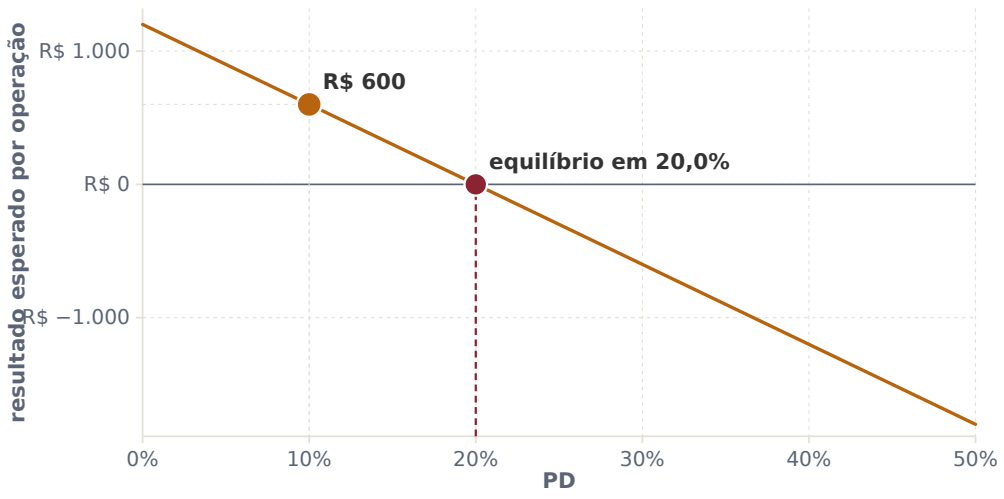


A mesma previsão pode sustentar políticas diferentes. A política é uma decisão adicional ao modelo.

Experimento sintético; semente e versões no notebook

Com estas hipóteses, o equilíbrio ocorre em PD de 20%

Aproximação didática por operação, horizonte de 12 meses



HIPÓTESES

aproximação didática, 12 meses

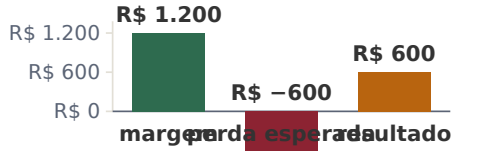
$$\mathbb{E}[\text{resultado}] = m - p\,\text{LGD}\,\text{EAD} \quad p^* = \frac{m}{\text{LGD}\,\text{EAD}} = 20\%$$

margem antes da perda	**R$ 1.200**
EAD	**R$ 10.000**
LGD	**60%**
perda esperada	**R$ 600**
resultado esperado	**R$ 600**

PD do contrato

R$ 600

MARGEM MENOS PERDA ESPERADA

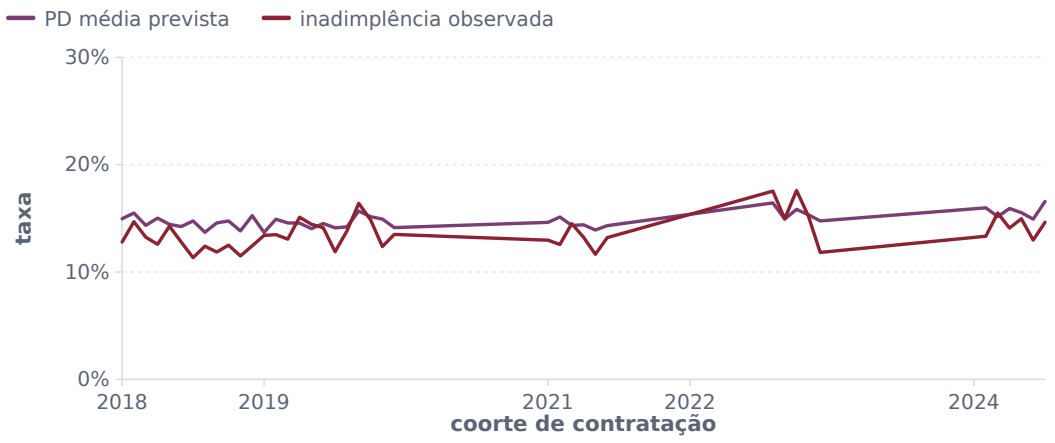


Um bom ranking só gera valor quando probabilidades e política se conectam à economia da operação.

Economia didática declarada na tela; cálculo próprio

O que conseguimos monitorar agora? O que exige esperar?

Indicadores imediatos e indicadores que dependem da maturação do alvo



Data de observação
jul/2025
41 coortes com alvo completo nesta data.

CENÁRIOS SINTÉTICOS

Os dois primeiros indicadores de cada cenário aparecem no dia seguinte à contratação. Inadimplência, calibração e discriminação só existem depois dos 12 meses de maturação.

INDICADORES IMEDIATOS

comprometimento médio da coorte mais recente	**35,5%**
utilização média	**45,8%**
ausência de utilização	**4,6%**
taxa de aprovação no corte congelado	**77,2%**
PD média prevista	**16,57%**

INDICADORES COM ALVO MATURADO

inadimplência da coorte mais recente já maturada	**14,63%**
coorte correspondente	**jul/2024**
coortes já maturadas	**41 de 41**

Podemos observar a carteira entrando hoje, mas parte da qualidade do modelo só será conhecida depois.

Experimento sintético; semente e versões no notebook

Defenda sua recomendação para o comitê

Desempenho, economia, estabilidade e capacidade de execução na mesma mesa

TESTE DE FEV A JUL DE 2024, 5.000 CONTRATOS E 712 EVENTOS

base sintética

EVIDÊNCIA NO TESTE	LOGIT	ÁRVORE	BOOSTING
AUC	0,7446	0,7392	0,7444
Brier, qualidade probabilística	0,10670	0,10823	0,10707
Taxa de aprovação	74,8%	77,7%	79,4%
Inadimplência entre aprovados	8,50%	9,32%	9,22%
Resultado por aprovado	R$ 690	R$ 641	R$ 647
Resultado total da carteira	R$ 2.578.800	R$ 2.487.600	R$ 2.565.600

Política congelada antes do teste: aprovar quando a PD calibrada não passa de 20%, que é o ponto de equilíbrio econômico do slide 47.

Modelo candidato Política

Restrição

Sem restrição adicional declarada.

ARGUMENTOS QUE A EVIDÊNCIA SUSTENTA

- A maior AUC no teste é de Logit, com 0,7446. A diferença para o menor é de 0,0054, pequena diante da variabilidade de uma única amostra.
- O maior resultado total sob a política escolhida é de Logit, com R$ 2.578.800 em 3.739 aprovados.
- Aprovação e inadimplência entre aprovados mudam junto com o modelo: comparar apenas AUC esconde a diferença de volume.
- Manter o logit como referência operacional e avaliar outro método como desafiante é uma recomendação defensável, se a evidência sustentar.

REQUISITOS DA OPERAÇÃO

- equipe enxuta, sem time dedicado a manutenção de modelos
- explicação da recusa exigida pelo canal e pela área de conduta
- acompanhamento mensal, com alvo maturando 12 meses depois

SUA JUSTIFICATIVA

Por que este modelo e esta política?

Qual é o principal risco desta escolha?

Como monitorar depois da implantação?

O texto fica apenas neste dispositivo.

Você consegue explicar, comparar e usar os três modelos?

Definir, estimar, validar e decidir são etapas conectadas

TÉCNICA	COMO CONSTRÓI A PD	PRINCIPAL CUIDADO	EVIDÊNCIA NECESSÁRIA
Regressão logística	soma funções das características em um escore e converte pela logística	especificação, unidade das variáveis e leitura dos coeficientes	validação da forma funcional e calibração fora do tempo
Árvore de decisão	aprende regras sucessivas e usa a frequência observada na folha	complexidade, tamanho de folha e instabilidade entre amostras	desempenho fora do tempo e estabilidade entre reamostragens
Gradient boosting	acrescenta contribuições sucessivas ao escore e converte no fim	trajetória de ajuste, complexidade de cada árvore e explicação	curva de validação, parada definida antes e verificação de calibração

Para os três: informação disponível na decisão, validação temporal, calibração das probabilidades e conexão com a decisão econômica.

OS MESMOS QUATRO CLIENTES DO COMEÇO

qual modelo, qual política e por quê?

Ana R$ 7.000 · comp 22% · hist não

Bruno R$ 4.500 · comp 38% · hist sim

Carla R$ 10.000 · comp 48% · hist não

Diego R$ 3.500 · comp 55% · hist sim

A técnica importa. A evidência que sustenta seu uso importa tanto quanto.

Síntese do curso

TRÊS PERGUNTAS DE RECUPERAÇÃO

Odds multiplicadas por 2 significam PD multiplicada por 2?
Não. Odds e probabilidade são escalas diferentes: com odds dobradas, 10% vira 18,18% e 40% vira 57,14%.

A mesma folha pode conter clientes diferentes?
Sim. A folha agrupa todos os contratos que satisfazem as mesmas regras, e atribui a mesma previsão a todos eles.

As árvores do boosting somam probabilidades?
Não. A soma acontece no escore F, e a transformação logística é aplicada uma única vez no final.

APLICAR NO MEU TRABALHO

1. Alvo: qual evento, em qual horizonte e para qual população elegível.
2. Dados: o que existia e estava acessível na data da decisão.
3. Referência: um modelo simples bem especificado, que os demais precisam superar.
4. Protocolo: partições por data, maturação do alvo e regras congeladas antes do teste.
5. Comparação: mesmo conjunto de informação, mesma métrica e incerteza declarada.
6. Decisão: corte, economia da operação e plano de monitoramento com responsável.

Leve os seis passos para a sua carteira. A conclusão da aula está na linha de fechamento, abaixo.

Apêndice: notas do professor e respostas

Mesma ordem dos slides. Cada bloco traz condução, respostas, cuidados e transição.

---

01 · Você aprovaria estes quatro clientes?

Esquema conceitual; sem dados

CONDUÇÃO EM AULA

- Peça 20 segundos de avaliação silenciosa e registre uma opinião por perfil.
- Pergunte por que Carla, com a maior renda da mesa, ainda poderia exigir investigação.
- Se alguém pedir taxa, prazo e custos, reconheça que faltam elementos da política econômica. A conta aparece no slide 47.

CUIDADOS E LIMITES

- A seleção representa apenas a escolha neste dispositivo. Não existe contagem de turma nem resultado coletivo.
- Nenhuma proposta tem desfecho conhecido. Não atribua valor moral à renda nem ao nome do cliente.
- Um cliente de maior risco ainda pode ser economicamente aceitável sob certas condições. Cobrar mais não resolve qualquer risco.

TRANSIÇÃO

Antes de escolher uma técnica, precisamos conhecer os mesmos clientes e definir exatamente o que vamos prever.

---

02 · O que sabemos no momento da proposta?

Esquema conceitual; sem dados

CONDUÇÃO EM AULA

- Compare primeiro a renda de Carla e a de Bruno. Depois revele comprometimento e utilização.
- Contraste relacionamento e histórico por último.
- Pergunte: qual combinação parece mais informativa que uma variável isolada? Aceite hipóteses diferentes.

CUIDADOS E LIMITES

- Canal não é efeito causal nem evidência de qualidade moral. Uma variável útil para previsão pode refletir seleção.
- Comprometimento é a parcela total mensal já incluindo a nova operação, dividida pela renda verificada.
- Histórico registra atraso de 15 a 89 dias no passado. Não é inadimplência corrente.
- As unidades serão preservadas em toda a aula, especialmente percentual contra fração.

TRANSIÇÃO

Conhecer os atributos não basta: o que significa exatamente risco nesta aula?

03 · Probabilidade de quê, para quem e até quando?

Trajetórias ilustrativas de contratos fictícios; cálculo próprio

CONDUÇÃO EM AULA

- Peça a classificação do contrato C antes de mostrar o status.
- Separe explicitamente ainda não aconteceu de não acontecerá dentro da janela.
- Leve o controle até o mês 12 e consolide os três elementos: população elegível, evento e horizonte.

RESPOSTAS

- A atinge 95 dias no mês 8, portanto y = 1.
- B fecha os 12 meses com máximo de 88 dias, portanto y = 0.
- C tem apenas 6 meses observados e nenhum evento: alvo incompleto, nunca y = 0.
- Chegar exatamente a 90 dias não satisfaz superior a 90 dias. No mês 7 o contrato A ainda não tem evento.

CUIDADOS E LIMITES

- Esta é uma definição operacional didática. Não é transcrição de norma bancária.
- Mudar o limite de atraso ou o horizonte muda o alvo previsto, e portanto muda o modelo.

TRANSIÇÃO

Agora sabemos o que prever. Precisamos garantir que os dados usados para prever existiam antes do resultado.

---

04 · Esse dado existia quando o crédito foi aprovado?

Registros didáticos de disponibilidade; cálculo próprio

CONDUÇÃO EM AULA

- Classifique primeiro as três fáceis: renda verificada, histórico anterior e utilização conhecida.
- Use o atraso máximo dos 12 meses seguintes para estabelecer a regra.
- Termine com a consulta tardia de bureau: a informação pode descrever algo anterior, mas o acesso e a versão precisam existir na decisão.

RESPOSTAS

- Renda verificada, histórico anterior, utilização conhecida na data e comprometimento com a parcela proposta estão disponíveis na decisão.
- Valor renegociado três meses depois, atraso máximo dos 12 meses seguintes e consulta de bureau posterior à inadimplência são informação do futuro.
- O comprometimento com a parcela proposta é permitido porque a parcela é conhecida na simulação, e não depende de observar o desfecho.

CUIDADOS E LIMITES

- Data de referência e data de disponibilidade são coisas diferentes.
- Limpeza, imputação, seleção de variáveis e normalização também vazam informação se aprenderem com o teste.

TRANSIÇÃO

Além de separar passado e futuro dentro de cada proposta, precisamos separar as amostras usadas para aprender e avaliar.

05 · Aprender no passado, escolher sem olhar o teste

Experimento sintético; semente e versões no notebook

CONDUÇÃO EM AULA

- Peça que o aluno aponte a diferença entre o fim das contratações e o fim da observação.
- Pergunte: podemos escolher o corte depois de ver os resultados finais? Revele que o teste avalia escolhas congeladas.
- Datas completas ficam aqui nas notas; a projeção mostra apenas os anos.

RESPOSTAS

- Treino: contratações de jan/2018 a dez/2019, alvo completo em dez/2020.
- Validação de ajuste: jan a jun/2021, alvo completo em jun/2022.
- Calibração e política: ago a dez/2022, alvo completo em dez/2023.
- Teste final: fev a jul/2024, alvo completo em jul/2025.

CUIDADOS E LIMITES

- Os espaços entre coortes tornam a maturação visível. Não são uma recomendação universal de desenho operacional.
- Divisão aleatória não é sempre inválida. Ela apenas não testa a generalização temporal buscada aqui.
- O modelo base é congelado antes da calibração, e a política antes do teste.

TRANSIÇÃO

Agora temos regras de comparação. Como cada técnica transforma os mesmos dados em uma PD?

---

06 · O mesmo cliente, três formas de aprender o risco

Esquema conceitual; sem dados

CONDUÇÃO EM AULA

- Passe rapidamente pelos três mecanismos, sem fórmulas.
- Pergunte qual parece mais fácil de explicar e se isso basta para escolher.
- Termine selecionando logit, que abre o próximo bloco.

CUIDADOS E LIMITES

- A árvore não é um conjunto de regras definidas pelo gestor: a estrutura é aprendida dos dados, embora o gestor imponha limites.
- Boosting não é votação majoritária nem soma de probabilidades.
- O logit não é incapaz de representar não linearidades.
- Nenhum painel traz PD de modelo: aqui só existe o mecanismo.

TRANSIÇÃO

Vamos começar pela ideia mais familiar: combinar características em um escore e converter esse escore em probabilidade.

07 · Cada informação altera o escore do cliente

Exemplo didático de logit; cálculo próprio

CONDUÇÃO EM AULA

- Comece no perfil de referência: renda R$ 5.000, comprometimento 30%, relacionamento 12 meses, utilização 40% e sem histórico de atraso.
- Acrescente comprometimento e histórico. Pergunte qual teve maior contribuição neste cliente.
- Mostre que um coeficiente pequeno pode gerar efeito relevante quando a variável muda muito.
- Revele z e pergunte se esse número já é uma probabilidade.

CUIDADOS E LIMITES

- A contribuição exibida é relativa à codificação de referência escolhida. Não é decomposição causal nem medida universal de importância.
- Coeficiente e contribuição observada são coisas diferentes: a contribuição multiplica o coeficiente pelo desvio da variável.
- Centrar as variáveis facilita a interpretação do intercepto e não muda a classe de funções.

APROFUNDAMENTO

- Material original, capítulo 4, página 13: o intercepto é âncora de nível, não coeficiente comum. Naquele exemplo, utilização zero e atraso zero não existem na carteira, então o intercepto não tem leitura como PD de um cliente real.
- Trocar o intercepto desloca a PD de todas as propostas e preserva a ordem de risco. É por isso que recalibrar um modelo cuja ordenação continua boa equivale a reestimar o nível, sem alterar a AUC. O mesmo fato aparece no slide 45.

TRANSIÇÃO

O escore pode ser qualquer número. Como converter esse escore em algo entre zero e um?

---

08 · −2,025 não é uma probabilidade

Exemplo didático de logit; cálculo próprio

CONDUÇÃO EM AULA

- Peça que o aluno calcule a reta simplificada em comprometimento de 20%.
- Revele o resultado negativo, leve o controle até 70% e mostre o excesso acima de 100%.
- Volte a Bruno e lembre que o escore dele não foi definido como probabilidade.

CUIDADOS E LIMITES

- A reta usada aqui existe para mostrar o problema. Não é modelo estimado.
- Não conclua que todo modelo linear de probabilidade é inútil. A mensagem é a dificuldade de garantir limites.
- Truncar em 0 e 1 altera a função de forma improvisada e não substitui um modelo probabilístico.

TRANSIÇÃO

A função logística faz exatamente essa conversão, preservando a ordem dos escores.

09 · Um escore, uma probabilidade entre 0% e 100%

Exemplo didático de logit; cálculo próprio

CONDUÇÃO EM AULA

- Comece em Bruno. Vá até z = 0 e pergunte o valor esperado antes de revelar.
- Compare o efeito de somar 1 ao escore em z = −5 e em z = 0.
- Termine em Ana e Bruno para mostrar que a função preserva a ordem dos escores.

CUIDADOS E LIMITES

- A sigmoide é estritamente crescente: a ordem de z e a ordem de p são a mesma.
- Um escore finito nunca produz exatamente 0 ou 1. A formatação pode arredondar para 0,00%, o que não significa impossibilidade.
- O ponto de 50% é referência matemática, não corte de aprovação.

APROFUNDAMENTO

- Material original, capítulo 4, página 6: probabilidade, odds, log odds e escore de bureau são quatro réguas da mesma ordem. Lá, PD de 5% corresponde a odds 0,053, log odds −2,944 e escore didático 865.
- Qualquer transformação monótona preserva a ordenação e, portanto, a AUC. O que muda é a régua de comunicação e de cálculo.

TRANSIÇÃO

Para interpretar os coeficientes, precisamos entender a escala em que a soma foi feita: o logaritmo das odds.

---

10 · 10% de PD significa odds de 1 para 9

Exemplo didático de logit; cálculo próprio

CONDUÇÃO EM AULA

- Pergunte 10 dividido por quê antes de revelar as odds.
- Mostre que trocar o denominador muda o conceito: 10 em 100 contra 10 para 90.
- Vá a 50% e depois a 20%. Termine lembrando que o escore z do logit é o log odds.

RESPOSTAS

- p = 50%: odds 1 e log odds 0.
- p = 20%: odds 0,25 e log odds aproximadamente −1,386.
- p = 10%: odds 1/9 aproximadamente 0,1111 e log odds aproximadamente −2,197.

CUIDADOS E LIMITES

- Odds não são outra porcentagem de risco. Podem passar de 1 e não se limitam a 100%.
- Uma razão escrita 1 para 9 é diferente do número decimal 0,1111 apenas na forma de escrever.
- Em p igual a 0 ou 1 o log odds vai a infinito. Por isso a entrada principal vai de 1% a 99%.

TRANSIÇÃO

Agora conseguimos interpretar o que acontece quando um coeficiente aumenta o log odds.

11 · Mais 10 pontos de comprometimento multiplicam as odds por 1,49

Exemplo didático de logit; cálculo próprio

CONDUÇÃO EM AULA

- Peça que os alunos antecipem a PD quando as odds dobram, começando pelo cenário de 10%.
- Revele primeiro 10%, depois 2% e 40%.
- Explique que a leitura vale mantidas as outras variáveis do modelo.

RESPOSTAS

- Com odds multiplicadas por 2: 2% vira aproximadamente 3,92%; 10% vira 18,18%; 40% vira 57,14%.
- Com mais 10 pontos de comprometimento, o multiplicador é e elevado a 0,40, aproximadamente 1,4918.

CUIDADOS E LIMITES

- Razão de chances de 1,49 não significa que o risco aumenta 49%.
- Para uma mudança finita use a conversão exata, não o efeito marginal local.
- A leitura é associativa e condicionada às demais variáveis. Não é efeito causal.
- Os três riscos iniciais são cenários didáticos. Não são os quatro clientes.

APROFUNDAMENTO

- Material original, capítulo 4, página 11: naquele exemplo, utilização a cada 10 pontos tem razão de chances 2,1071 e atraso a cada 10 dias tem 4,0370. A leitura está correta como multiplicação das chances e não autoriza leitura causal.
- Página 14 do mesmo capítulo: trocar a unidade muda o coeficiente e não muda o modelo. A utilização vale 0,74530 por dezena de pontos, 0,07453 por ponto e 7,45300 por fração do limite, com PD prevista idêntica nos três casos.
- Disso decorre que comparar 1,3955 com 0,7453 para dizer qual variável pesa mais não significa nada: as unidades são diferentes. A comparação legítima usa um movimento realista de cada variável, como um desvio padrão, que é o que o slide 18 faz com coeficientes padronizados.

TRANSIÇÃO

Vamos aplicar essa interpretação a um cliente completo e conferir todas as etapas da conta.

---

12 · Bruno: como chegamos à PD de 11,66%?

Exemplo didático de logit; cálculo próprio

CONDUÇÃO EM AULA

- Reconstrua a conta passo a passo antes de mexer nos controles.
- Peça uma alteração que reduza a PD e outra que a aumente.
- Questione se a alteração seria factível: reduzir comprometimento sem mudar parcela ou renda é uma simulação abstrata.

RESPOSTAS

- Ana 1,10%, Bruno 11,66%, Carla 2,82% e Diego 27,59%, com a fórmula manual.
- O recurso descreve a sensibilidade do modelo, não a garantia de uma ação causal disponível ao cliente.

CUIDADOS E LIMITES

- Canal aparece na ficha e não participa desta fórmula manual.
- Esta PD é do exemplo didático. Não é previsão de modelo treinado.

TRANSIÇÃO

Esses coeficientes foram escolhidos para tornar a conta transparente. Como um modelo aprende coeficientes a partir dos dados?

13 · O modelo aprende penalizando previsões incompatíveis com os dados

Exemplo didático de logit; cálculo próprio

CONDUÇÃO EM AULA

- Escolha y = 1 com p = 10% e depois aumente p até 90%.
- Troque y para 0 mantendo p = 90% e pergunte por que a penalização aumenta.
- Mostre que uma observação não define sozinha os coeficientes: a perda total agrega todos os casos.

RESPOSTAS

- Para y = 1: p = 0,10 gera perda aproximadamente 2,3026; p = 0,90 gera aproximadamente 0,1054.
- Para y = 0 a relação se inverte: p = 0,90 gera aproximadamente 2,3026.
- Minimizar a log loss equivale a maximizar a verossimilhança do modelo logístico.

CUIDADOS E LIMITES

- Classificar corretamente a partir de um corte não basta para medir a qualidade da probabilidade.
- A perda depende do grau de confiança, não apenas do lado do corte.
- A função objetivo do treino não garante o melhor desempenho futuro.
- Em p igual a 0 ou 1 a perda vai a infinito. As curvas são desenhadas entre 0,1% e 99,9%.

APROFUNDAMENTO

- Material original, capítulo 4, páginas 15 a 17: a perda de cada observação determina os coeficientes, e a descida de gradiente percorre os parâmetros uma iteração por vez até a convergência.
- Nesta aula a descida aparece no bloco de boosting, slides 32 a 35, com a mesma mecânica de resíduo, passo e atualização do escore.
- Página 18 do original: navegador, Python com a mesma regra e sklearn chegam ao mesmo resultado, com diferença máxima de 0,000327 no intercepto. A conferência equivalente aqui está nos metadados do slide 43 e no notebook.

TRANSIÇÃO

Mesmo com um coeficiente fixo, o efeito observado na PD muda ao longo da curva logística.

---

14 · Aumentar o escore em 0,40 não aumenta toda PD pelo mesmo valor

Exemplo didático de logit; cálculo próprio

CONDUÇÃO EM AULA

- Comece pelos três segmentos, todos com a mesma largura horizontal, e peça que comparem as alturas.
- Revele os valores e só então a derivada.
- Reduza a mudança para 1 ponto e mostre a aproximação local ficando mais próxima.

RESPOSTAS

- O efeito máximo da derivada da logística ocorre em p igual a 0,5. Isso é matemática, não recomendação de operar perto de 50% de PD.
- A derivada é aproximação para mudanças pequenas. Para 10 pontos de comprometimento, use a conversão exata.

CUIDADOS E LIMITES

- A variação de z continua sendo o coeficiente vezes a variação da variável.
- Valores em pontos percentuais não podem ser lidos em unidade de odds.
- Em variação zero, os dois métodos dão exatamente zero.

TRANSIÇÃO

Até aqui usamos variáveis numéricas e um indicador. Como incluir uma característica com várias categorias?

15 · Canal digital não significa duas vezes agência

Extensão pedagógica da fórmula manual; cálculo próprio

CONDUÇÃO EM AULA

- Pergunte o que ficaria implícito ao usar canal igual a 1, 2 e 3 diretamente.
- Mostre a tabela de indicadores e só então troque a referência.
- Use o momento para separar interpretação de coeficientes e previsão do modelo.

RESPOSTAS

- Com agência como referência: digital soma 0,30 e parceiro soma 0,60 ao escore.
- Com digital como referência: intercepto passa a −3,20, agência recebe −0,30 e parceiro +0,30. As PDs por categoria não mudam.
- Categoria omitida não é categoria ausente: ela está no intercepto.

CUIDADOS E LIMITES

- Este é um exemplo separado, que não altera a fórmula canônica dos slides 07 a 14 nem as PDs dos quatro clientes.
- Uma categoria nova em produção exige tratamento definido na codificação. Não invente coeficiente.
- Canal pode refletir seleção, não causalidade.
- Manter todas as indicadoras junto com o intercepto cria redundância na parametrização.

TRANSIÇÃO

A codificação determina a forma que o modelo enxerga a informação. Podemos também flexibilizar a forma das variáveis numéricas.

---

16 · O efeito pode se intensificar depois de um limite

Cenário separado da fórmula manual; cálculo próprio

CONDUÇÃO EM AULA

- Mostre primeiro a relação na escala do escore.
- Ative a transformação e peça que identifiquem onde a inclinação muda.
- Acompanhe a tradução para PD e pergunte como decidir se a flexibilidade ajudou: a resposta cita validação, não a beleza da curva.

RESPOSTAS

- Antes de 40% a inclinação é 0,04 por ponto. Depois de 40% passa a 0,08 por ponto.
- A função é contínua no nó: em comprometimento de 40% as duas formas coincidem.

CUIDADOS E LIMITES

- Escolher o nó olhando o teste vaza informação.
- Flexibilidade custa parâmetros e exige controle de complexidade.
- Este é um exemplo linear por partes, não uma spline cúbica.
- Árvores não são necessárias para representar toda não linearidade.

APROFUNDAMENTO

- Material original, capítulo 4, página 21: faixas devolvem flexibilidade sem trocar de família, ao custo de um coeficiente por faixa em vez de um por variável.
- As regras declaradas lá continuam válidas: volume mínimo por faixa, monotonicidade quando existe hipótese econômica de direção, estabilidade das fronteiras nas safras seguintes e fronteiras definidas antes de olhar o alvo.
- O WoE é esse mesmo tratamento com um valor específico por faixa, o logaritmo da razão entre a distribuição dos adimplentes e a dos inadimplentes. A mudança de inclinação deste slide é a alternativa contínua, com um parâmetro adicional apenas.

TRANSIÇÃO

Além de mudar a forma de um efeito, podemos permitir que ele dependa de outra característica.

17 · O comprometimento pesa igual para quem tem histórico de atraso?

Cenário separado da fórmula manual; cálculo próprio

CONDUÇÃO EM AULA

- Comece pelo escore, com as duas retas paralelas.
- Pergunte o que muda quando o termo cruzado é ativado. Só então observe a PD.
- Peça que o aluno separe efeito do histórico no escore de diferença de probabilidade.

RESPOSTAS

- Sem interação: inclinação 0,04 nos dois grupos, retas paralelas no escore.
- Com interação de 0,03: inclinação 0,04 para histórico zero e 0,07 para histórico um.
- Mesmo sem termo cruzado, a diferença de PD entre os grupos varia com o comprometimento, porque a sigmoide é não linear.

CUIDADOS E LIMITES

- Os coeficientes principais são condicionais ao valor de referência das demais variáveis.
- Com interação, o coeficiente do histórico não é efeito constante em qualquer comprometimento.
- A leitura é associativa e pode não representar uma intervenção viável.
- Ausência de termo cruzado não implica diferença constante de PD.

TRANSIÇÃO

Cada transformação ou interação acrescenta flexibilidade. Como evitar que isso capture particularidades da amostra?

---

18 · Coeficientes extremos podem estar aprendendo ruído

Experimento sintético; semente e versões no notebook

CONDUÇÃO EM AULA

- Comece com penalização fraca, observe as magnitudes e a validação.
- Aumente a penalização e discuta o que melhora e o que piora.
- Compare os dois cenários: na amostra completa o efeito é pequeno; na amostra reduzida ele aparece com clareza.

CUIDADOS E LIMITES

- Regularização reduz variância em muitas situações. Não garante estabilidade em qualquer conjunto nem resolve mudança de população.
- Coeficientes padronizados não são os números da fórmula manual dos slides 07 a 14.
- A escolha da penalização faz parte do treinamento, e usa validação, nunca o teste.
- Com variáveis correlacionadas, nem todo coeficiente diminui de magnitude de forma monótona.
- A força da penalização mostrada é 1 dividido por C. C menor significa penalização maior.

TRANSIÇÃO

Podemos tornar o logit mais flexível e controlar sua complexidade. Em quais situações essa combinação é especialmente útil?

19 · Explicar a estrutura é vantagem; especificar a estrutura é responsabilidade

Síntese conceitual aplicada; sem dados

CONDUÇÃO EM AULA

- Peça uma razão para manter o logit mesmo se outro algoritmo tiver AUC um pouco maior.
- Aceite manutenção, clareza de especificação e estabilidade, sempre condicionadas a evidência.
- Depois peça uma situação em que insistir em um logit mal especificado seria prejudicial.

CUIDADOS E LIMITES

- Interpretável depende da quantidade de termos, transformações, interações e do público. Um logit com centenas de variáveis não é automaticamente fácil de explicar.
- Coeficiente não é causa.
- O desempenho de uma formulação simples não é o limite de toda a classe.
- Produzir uma probabilidade não garante que ela esteja calibrada.

APROFUNDAMENTO

- Material original, capítulo 4, página 20: a linearidade em log odds é o limite da família. Um padrão não monótono na utilização não é representável sem transformar a variável.
- Página 19: a fronteira de decisão do logit é uma reta no plano das variáveis, porque o escore é uma soma. Mudar o corte desloca a reta paralelamente e mudar um coeficiente gira a reta. Compare com as regiões retangulares da árvore no slide 26.
- Página 22: a síntese do capítulo original é soma, curva e limite. Os três aparecem aqui nos slides 07, 09 e neste.

TRANSIÇÃO

Antes de mudar de técnica, vamos verificar se conseguimos usar o que aprendemos em uma decisão de interpretação.

---

20 · Bruno aumenta seu comprometimento de 38% para 48%. E agora?

Exemplo didático de logit; cálculo próprio

CONDUÇÃO EM AULA

- Dê um minuto em duplas. Peça a explicação de quem escolheu cada alternativa, sem expor pessoas.
- Use a solução para revisar a diferença entre coeficiente e alteração de PD.
- Pergunte oralmente: o que podemos afirmar sobre o efeito causal de reduzir o comprometimento?

RESPOSTAS

- A alternativa correta é a segunda: o escore sobe 0,40 e as odds são multiplicadas por e elevado a 0,40.
- Escore original −2,025 e PD 11,66%. Novo escore −1,625 e PD aproximadamente 16,45%. Diferença aproximada de 4,79 pontos.
- Sobre causalidade: o cálculo descreve o modelo mantendo as demais entradas fixas. Não identifica efeito causal.
- Desafio opcional: sem o histórico de atraso, o escore de Bruno cai 0,80 e a PD passa de 11,66% para aproximadamente 5,66%.

CUIDADOS E LIMITES

- O que dobra são as odds, não a probabilidade. Converter odds em PD exige a volta pela logística.
- Os coeficientes deste exemplo foram escolhidos para ensinar, não estimados em dados. A ordem de grandeza é plausível; os valores não são um modelo.
- Uma PD de Bruno não diz o que acontecerá com Bruno. Diz a frequência esperada em exposições semelhantes.

TRANSIÇÃO

Agora imagine que o modelo não some características, mas faça perguntas sucessivas: houve atraso? O comprometimento é alto? Essa é a lógica de uma árvore.

21 · Houve atraso? Qual é o comprometimento?

Árvore didática de 1.000 contratos; cálculo próprio

CONDUÇÃO EM AULA

- Pergunte qual informação a árvore consultará primeiro.
- Leve Ana à esquerda e então à folha de comprometimento até 40%.
- Mostre Diego no outro ramo para criar contraste. Só depois nomeie nó, ramo e folha.
- Não explique Gini neste slide.

RESPOSTAS

- Ana: histórico 0 e comprometimento 22%, portanto ramo esquerdo e folha de comprometimento até 40%.
- Bruno: histórico 1 e comprometimento 38%, portanto ramo direito e folha de comprometimento até 40%.
- Carla: histórico 0 e comprometimento 48%, folha de comprometimento acima de 40%.
- Diego: histórico 1 e comprometimento 55%, folha de comprometimento acima de 40%.
- Comprometimento exatamente igual a 40 vai para o ramo até 40.

CUIDADOS E LIMITES

- O percurso é determinístico, mas a previsão ao final é probabilística. Duas pessoas diferentes podem cair na mesma folha.
- As perguntas e os limites são aprendidos dos dados. O limite de 40% é regra deste exemplo, não norma de crédito.
- Esta árvore pequena serve para enxergar o mecanismo. Não é o resultado do experimento maior.

TRANSIÇÃO

Depois de encaminhar o cliente, precisamos atribuir uma PD ao grupo ao qual ele pertence.

---

22 · Nesta folha, 18 de 600 contratos ficaram inadimplentes

Árvore didática de 1.000 contratos; cálculo próprio

CONDUÇÃO EM AULA

- Construa a conta de Ana em voz alta e peça a de Bruno.
- Pergunte se clientes da mesma folha terão a mesma previsão mesmo com rendas diferentes: sim, nesta árvore, porque a renda não entra nas perguntas.

RESPOSTAS

- Ana 18/600 = 3%. Bruno 12/80 = 15%. Carla 22/200 = 11%. Diego 48/120 = 40%.
- As folhas somam 1.000 contratos e 100 eventos, iguais à raiz.

CUIDADOS E LIMITES

- A taxa aparente da folha não comprova calibração futura.
- Este exemplo é sem pesos. Com ajuste ponderado, a fórmula da folha refletiria os pesos.
- Estas PDs são da árvore didática. Não são previsões do experimento estimado, e não devem ser comparadas com as PDs manuais do logit como se uma fosse mais correta.

TRANSIÇÃO

Por que a primeira pergunta foi sobre histórico, e não sobre renda ou comprometimento?

23 · Qual pergunta separa melhor os comportamentos observados?

Árvore didática de 1.000 contratos; cálculo próprio

CONDUÇÃO EM AULA

- Peça que o aluno escolha visualmente e explique por quê.
- Se ele usar apenas a diferença entre taxas, pergunte o que aconteceria com uma folha de duas observações.
- Mostre o papel do tamanho do grupo antes de revelar a regra de escolha.

RESPOSTAS

- Histórico separa em 800 contratos com 5% e 200 contratos com 30%.
- Comprometimento separa em 680 contratos com aproximadamente 4,41% e 320 contratos com 21,88%.
- A regra de escolha é a menor impureza ponderada, calculada no próximo slide.

CUIDADOS E LIMITES

- Esta é a melhor divisão entre os candidatos comparados, não entre todas as variáveis possíveis.
- A busca é gulosa e local: a melhor divisão imediata não garante a árvore globalmente ótima.
- Com dados agregados não é possível varrer limiares intermediários sem inventar contagens.

TRANSIÇÃO

Vamos colocar essa comparação em números com o índice de Gini.

---

24 · Histórico reduz o Gini de 0,180 para 0,160

Árvore didática de 1.000 contratos; cálculo próprio

CONDUÇÃO EM AULA

- Peça a conta do Gini da raiz antes de revelar.
- Antes de ponderar, pergunte por que não usar a média simples. Mostre a diferença entre 800 e 200 registros.
- Termine explicitando que o algoritmo compara vários candidatos, repete o procedimento em cada nó e precisa de uma regra para parar.

RESPOSTAS

- Raiz com 10% de eventos: Gini igual a 0,180.
- Por histórico: 0,8 vezes 0,095 mais 0,2 vezes 0,420 resulta em 0,160, com ganho de 0,020.
- Por comprometimento: impureza ponderada 0,166728 e ganho 0,013272.
- Impureza zero em grupo puro; máximo de 0,5 quando a taxa é 50%.

CUIDADOS E LIMITES

- Gini é critério de impureza no treino. Não é AUC, KS nem lucro.
- Outras perdas podem orientar árvores. Não misture entropia e Gini no mesmo cálculo.
- Ganho de impureza não é efeito causal do histórico.

TRANSIÇÃO

Ao dividir mais, as folhas parecem mais precisas. Mas quanto podemos confiar em uma taxa calculada com poucas observações?

25 · Duas folhas com PD de 10%. A evidência é igual?

Pares binomiais ilustrativos; cálculo próprio

CONDUÇÃO EM AULA

- Pergunte qual folha transmitiria mais confiança e por quê.
- Aumente o tamanho da folha e peça que o aluno narre o que mudou.
- Mostre o caso de zero eventos e pergunte se o gestor poderia tratar o risco como impossível.

RESPOSTAS

- 2 em 20: intervalo aproximado de 2,8% a 30,1%.
- 100 em 1.000: intervalo aproximado de 8,3% a 12,0%.
- 0 em 20: o limite superior continua positivo, aproximadamente 16,1%. Zero evento não prova risco zero.

CUIDADOS E LIMITES

- O intervalo descreve a dimensão binomial. Não incorpora a escolha adaptativa das folhas, a seleção da população nem a mudança temporal.
- Não é correto dizer que a PD individual tem 95% de chance de cair nesse intervalo.
- A incerteza real da árvore inclui a busca pelos cortes, e é maior que a mostrada aqui.

APROFUNDAMENTO

- Fórmula do intervalo de Wilson, com z igual a 1,96 para 95% nominal: o centro é $(\hat{p} + z^2/2n)$ dividido por $(1 + z^2/n)$, e a meia amplitude é z vezes a raiz de $[\hat{p}(1-\hat{p})/n + z^2/4n^2]$, dividida pelo mesmo denominador.
- Wilson é preferível ao intervalo normal simples justamente no caso deste slide: taxa baixa e folha pequena, onde o intervalo normal produz limite inferior negativo.

TRANSIÇÃO

Folhas maiores estabilizam estimativas, mas a árvore também precisa capturar diferenças relevantes entre clientes.

---

26 · O mesmo aumento de comprometimento pode mudar de folha

Árvore didática de 1.000 contratos; cálculo próprio

CONDUÇÃO EM AULA

- Mova Bruno de 38 para 41 e compare com Ana ao redor do mesmo limite.
- A mudança é de 15% para 40% no ramo com histórico, e de 3% para 11% no ramo sem histórico.
- Pergunte se uma mudança pequena de risco verdadeiro deveria produzir esse salto: a descontinuidade é da aproximação do modelo.

CUIDADOS E LIMITES

- Árvores capturam interações porque a pergunta de um ramo depende das decisões anteriores. Isso não dispensa validação.
- Uma folha define uma região constante. Ensembles produzem funções mais detalhadas.
- O limite de 40% é deste exemplo. Não é política de crédito.
- Histórico é categórico: as duas faixas são categorias, não um eixo contínuo.

TRANSIÇÃO

Se continuarmos dividindo o espaço, podemos criar regiões cada vez menores. Quando isso começa a prejudicar?

27 · A árvore cresce. O resultado fora do treino acompanha?

Experimento sintético; semente e versões no notebook

CONDUÇÃO EM AULA

- Comece raso, aumente a complexidade e peça que comparem treino e validação.
- Mostre a mediana de contratos por folha como pista de fragilidade, não como prova isolada.
- A seleção usa a validação fora do tempo, nunca o teste final.

CUIDADOS E LIMITES

- Os mapas vêm de um experimento auxiliar com apenas duas variáveis. Não são projeção exata do modelo completo.
- A perda mostrada usa log loss com proteção numérica de 10 elevado a menos 6, porque árvores produzem probabilidades iguais a 0 ou 1 em folhas puras. O Brier está disponível na tabela.
- Desempenho de uma única amostra varia. Diferenças pequenas podem não ser robustas.
- Nem toda árvore profunda é ruim: a profundidade interage com tamanho de folha e poda.

TRANSIÇÃO

Podemos controlar o crescimento de mais de uma maneira. Cada parâmetro restringe uma parte diferente da árvore.

---

28 · Três controles, três formas de conter o excesso de divisões

Experimento sintético; semente e versões no notebook

CONDUÇÃO EM AULA

- Mostre primeiro a profundidade, depois uma folha pequena que ainda existe mesmo em árvore rasa.
- Use essa folha para motivar o mínimo de contratos por folha.
- Por fim, explique que a poda considera retirar partes de uma árvore já crescida.

RESPOSTAS

- O parâmetro de folha é mínimo de observações, não mínimo de inadimplentes.
- O alfa da poda não tem unidade universal de nós removidos: ele é o preço por folha no critério de custo e complexidade.
- A seleção usa a perda de validação fora do tempo.

CUIDADOS E LIMITES

- Cada configuração corresponde a uma árvore realmente treinada, não a uma ocultação visual de nós.
- Aumentar o mínimo por folha não melhora calibração automaticamente.
- O critério final continua sendo desempenho fora da amostra de ajuste.

TRANSIÇÃO

Mesmo árvores de tamanho parecido podem mudar bastante quando mudamos a amostra usada para aprender.

---

29 · Se mudarmos um pouco a amostra, a árvore continua contando a mesma história?

Experimento sintético; semente e versões no notebook

CONDUÇÃO EM AULA

- Pergunte qual das duas árvores é a verdadeira e use a pergunta para discutir estimativa e variabilidade.
- Compare estrutura e previsão: uma raiz diferente não significa mudança enorme de PD, e previsões podem variar mesmo com raízes iguais.

CUIDADOS E LIMITES

- O bootstrap ilustra sensibilidade ao treinamento. Não é intervalo preditivo calibrado do risco verdadeiro de um cliente.
- Não reamostramos o teste para treinar.
- As réplicas exibidas foram escolhidas por critério declarado: a primeira réplica de cada variável de raiz observada, e não a mais dramática.
- Este slide não é uma aula de random forest. Apenas situa por que ensembles exploram múltiplas funções.

TRANSIÇÃO

Antes de combinar várias árvores, vamos verificar se sabemos interpretar uma única árvore e suas limitações.

30 · Carla recebe qual PD nesta árvore?

Árvore didática de 1.000 contratos; cálculo próprio

CONDUÇÃO EM AULA

- Reserve dois minutos em duplas.
- Não revele a resposta ao primeiro clique em um ramo errado: explique a regra e permita corrigir.
- Conclua com a importância de tamanho de folha, validação e estabilidade. Este é um bom ponto para a pausa da aula.

RESPOSTAS

- Carla tem histórico 0 e comprometimento 48%, acima de 40%. Chega à folha de 22 eventos em 200 contratos, ou seja 11%.
- Dobrar a renda isoladamente, com o comprometimento mantido fixo, não altera esta árvore, porque a renda não é usada nas perguntas.
- Numa mudança econômica real o comprometimento poderia variar. A simulação é sobre a função do modelo.
- Desafio final: Bruno de 38% para 41% passa de 15% para 40%. Isso não prova um salto real de risco exatamente nesse limite: é uma descontinuidade da aproximação aprendida.

CUIDADOS E LIMITES

- A taxa de uma folha é estimativa com incerteza, e folhas pequenas têm intervalo largo. Ler a taxa sem o tamanho da folha é o erro mais comum aqui.
- A regra da árvore é associação aprendida na amostra, não relação causal: mudar a variável no cliente não move a PD dele pelo caminho da árvore.
- Esta árvore vem da miniatura de 1.000 contratos, separada da base do experimento. Não compare o desempenho dela com o dos modelos treinados.

TRANSIÇÃO

Uma única árvore é clara, mas pode ser limitada ou instável. Podemos construir uma previsão melhor acrescentando várias árvores pequenas em sequência?

---

31 · E se uma árvore pequena ainda deixar estrutura nos erros?

Experimento auxiliar de duas variáveis; cálculo próprio

CONDUÇÃO EM AULA

- Retome o slide 27: aumentar uma única árvore não é a única opção.
- Mostre que há grupos em que a previsão ficou sistematicamente baixa ou alta.
- Pergunte como uma próxima função poderia atuar nesses grupos. Termine apresentando a palavra sequencial.

CUIDADOS E LIMITES

- Um resíduo individual não é padrão. A leitura só faz sentido na média por região.
- Subajuste e variabilidade não se resolvem com boosting em todos os casos: combinar árvores também acrescenta complexidade e pode sobreajustar.
- Os resíduos aqui vêm do conjunto usado no ajuste, nunca do teste.
- O título é uma hipótese a ilustrar com dados, não um diagnóstico automático.

TRANSIÇÃO

Vamos entender a regra de construção: partir de um escore simples, calcular a direção de melhoria e acrescentar uma árvore.

32 · A próxima árvore aprende a partir das previsões atuais

Miniatura didática de 10 registros; cálculo próprio

CONDUÇÃO EM AULA

- Pergunte se cada árvore produz uma nova PD a ser somada.
- Revele que o objeto somado é F, um escore na escala de log odds neste caso.
- Reconecte ao slide 09: só ao final aplicamos a função logística.

RESPOSTAS

- Novo escore é igual ao escore atual mais a taxa de aprendizagem vezes a contribuição da árvore.
- A árvore de regressão devolve números positivos ou negativos, não um voto entre adimplente e inadimplente.
- Em combinação independente, os modelos são ajustados sem esse encadeamento e depois agregados.

CUIDADOS E LIMITES

- O algoritmo pode otimizar perdas diferentes. Aqui o alvo é binário e a perda é logística.
- O gradiente depende da função de perda escolhida.
- Esta é uma apresentação de primeira ordem, didaticamente válida. Implementações podem usar atualizações diferentes nas folhas.

TRANSIÇÃO

Antes de corrigir qualquer coisa, precisamos de uma previsão inicial. Qual seria a melhor se tratássemos todos igualmente?

---

33 · Dois eventos em dez contratos: começamos em 20%

Miniatura didática de 10 registros; cálculo próprio

CONDUÇÃO EM AULA

- Pergunte qual PD seria razoável antes de olhar qualquer característica.
- Compare 10%, 20% e 40% na perda média.
- Revele que a média empírica minimiza a perda do modelo constante não ponderado, e conecte F0 à conversão do slide 10.
- Só então mostre os grupos A e B, anunciando que eles permitirão melhorar a previsão uniforme.

RESPOSTAS

- Média de y igual a 0,2, que é o mínimo da curva de perda.
- $F0 = \ln(0{,}20 / 0{,}80) = -1{,}3863$, com logaritmo natural.

CUIDADOS E LIMITES

- O ponto inicial depende da perda, dos pesos e da implementação. Este exemplo é sem pesos e binário.
- A taxa de 20% descreve esta miniatura. Não descreve o mercado de crédito nem a base principal.
- Os desfechos aqui são atributos de treinamento conhecidos, não informação sobre um cliente novo.
- Ao avançar para a miniatura canônica, a previsão inicial volta a 20%.

TRANSIÇÃO

Agora compare o que aconteceu com o que previmos: essa diferença indica a direção da primeira correção.

34 · A primeira correção reduz A e aumenta B

Miniatura didática de 10 registros; cálculo próprio

CONDUÇÃO EM AULA

- Peça o resíduo de um evento e o de um não evento.
- Calcule a média do grupo B em voz alta.
- Pergunte por que três adimplentes do grupo B receberam aumento. Resposta: o modelo captura o risco do grupo, não reproduz cada rótulo.

RESPOSTAS

- Em A, cinco resíduos de −0,2 e média −0,2.
- Em B, três resíduos de −0,2 e dois de +0,8, média +0,2.
- Com taxa de aprendizagem 1: F do grupo A vai a −1,586294 e PD a 16,9906%. F do grupo B vai a −1,186294 e PD a 23,3922%.

CUIDADOS E LIMITES

- y menos p é o gradiente negativo da perda logística em relação ao escore nesta parametrização.
- Isso não significa que toda biblioteca use exatamente a média como valor final da folha.
- As contribuições estão em unidades de escore, não em pontos percentuais de PD.
- Primeira ordem didática: a folha recebe a média do gradiente negativo, sem passo de Newton.

TRANSIÇÃO

Depois da atualização, os resíduos mudaram. Por isso a segunda árvore precisa olhar novamente para as previsões.

---

35 · A primeira árvore mudou as previsões. Agora mudam os gradientes

Miniatura didática de 10 registros; cálculo próprio

CONDUÇÃO EM AULA

- Antes de revelar a média do grupo B, pergunte se ela continuará em mais 0,2.
- Faça a turma perceber que o resíduo de um y = 1 diminuiu e o de um y = 0 ficou mais negativo.
- Mostre que a correção média positiva persiste, mas mudou de tamanho.

RESPOSTAS

- No grupo A, r1 = −0,1699058933 e a média da folha é o mesmo valor.
- No grupo B, três resíduos de −0,2339223413 e dois de +0,7660776587, média +0,1660776587.
- Com taxa 1: F do grupo A vai a −1,7562002544 e PD a 14,7267%. F do grupo B vai a −1,0202167024 e PD a 26,4985%.

CUIDADOS E LIMITES

- Os resíduos da segunda rodada usam p1, nunca p0.
- A nova árvore poderia usar outro corte em dados mais ricos. Aqui o corte permanece por construção da miniatura, que só tem dois valores de comprometimento.
- Nem toda correção diminui, e as previsões individuais não caminham necessariamente de forma monótona. Isso acontece aqui pela simplicidade dos grupos.
- O objetivo é reduzir a perda agregada, não acertar cada rótulo.

TRANSIÇÃO

Somamos duas contribuições. Vamos olhar essa soma como uma função que também pode prever um cliente novo.

36 · Escore inicial mais árvore 1 mais árvore 2 é a nova PD

Miniatura didática de 10 registros; cálculo próprio

CONDUÇÃO EM AULA

- Apresente o solicitante novo e pergunte onde está o y dele.
- Mostre que y só foi usado no treinamento das duas árvores.
- Faça o percurso nas duas árvores, some e converta. Compare com o waterfall do logit: ambos somam em uma escala de escore, com funções diferentes.

RESPOSTAS

- Grupo B: −1,386294 mais 0,200000 mais 0,166078 é igual a −1,020217, e a PD é 26,4985%.
- Grupo A: −1,386294 menos 0,200000 menos 0,169906 é igual a −1,756200, e a PD é 14,7267%.

CUIDADOS E LIMITES

- Não some as PDs de cada árvore nem chame as contribuições de probabilidades.
- Um conjunto real tem muitas árvores e padrões mais ricos. A miniatura de duas árvores só torna o mecanismo visível.
- Esta não é uma explicação por variável: é a decomposição exata da soma das árvores.
- Com taxa de aprendizagem diferente de 1, o que se soma é a taxa multiplicada pela contribuição.

TRANSIÇÃO

Se cada passo for menor, precisaremos de mais etapas. Como escolher o tamanho dos passos e o número de árvores?

---

37 · Learning rate e número de árvores precisam ser escolhidos juntos

Experimento sintético; semente e versões no notebook

CONDUÇÃO EM AULA

- Compare as taxas no mesmo número de árvores e depois permita mais árvores para a taxa menor.
- Pergunte se taxa menor é sempre melhor. Não: há custo computacional, subajuste com poucas árvores e comportamento que depende dos dados.

RESPOSTAS

- Cada curva vem de um treinamento próprio com sua taxa. Não é a reescala de um conjunto fixo de árvores.
- Multiplicar as contribuições no fim não reproduz outro treinamento, porque os resíduos dependem do escore atualizado a cada passo.

CUIDADOS E LIMITES

- O produto entre taxa e número de árvores não determina sozinho o modelo final: a trajetória muda.
- A taxa se relaciona com regularização, mas não substitui validação temporal.
- O teste final continua fora de qualquer escolha.

TRANSIÇÃO

Além de quantos passos dar e do tamanho de cada passo, importa o que cada árvore consegue representar.

---

38 · Uma correção pode depender de uma característica ou da combinação de várias

Experimento auxiliar de duas variáveis; cálculo próprio

CONDUÇÃO EM AULA

- Mostre a combinação de comprometimento alto com utilização alta e pergunte que regra a descreve.
- Compare a representação por uma única condição e por duas condições.
- Revele que árvores mais profundas capturam estruturas mais complexas, e que cada passo também pode ajustar mais ruído.

CUIDADOS E LIMITES

- Profundidade máxima não é contagem universal e exata da ordem de interações. Profundidade 2 permite caminhos com duas condições, e implementações e repetição de variável importam.
- A soma de stumps é aditiva no escore F. A transformação logística pode induzir não aditividade na escala da probabilidade. Não diga que stumps não produzem interação na PD.
- Este experimento auxiliar não fala sobre o desempenho na base de crédito principal.

TRANSIÇÃO

Temos vários controles de complexidade. O acompanhamento da validação ajuda a decidir quando interromper a construção.

39 · Continuar melhorando o treino pode deixar de ajudar a validação

Experimento sintético; semente e versões no notebook

CONDUÇÃO EM AULA

- Peça que a turma sugira onde parar antes de revelar toda a curva.
- Compare o palpite com a regra: tolerância e paciência evitam reagir a variações pequenas.
- Mostre que o mínimo aparente pode ser ruidoso.

CUIDADOS E LIMITES

- Melhor iteração e iteração de parada são coisas diferentes. A parada acontece depois da paciência, não no mínimo.
- A validação usada é fora do tempo, definida no protocolo. Não é um corte aleatório interno da biblioteca.
- Escolher o número de árvores é seleção de hiperparâmetro, e usa a mesma amostra de validação.
- Um eventual reajuste com mais dados precisa respeitar a disponibilidade temporal e ser definido antes de calibrar e testar.

TRANSIÇÃO

Com esses controles, podemos avaliar em que condições a flexibilidade do boosting oferece ganhos e quais custos ela traz.

---

40 · Mais flexibilidade exige mais disciplina de validação

Síntese conceitual, com um resultado calculado como exemplo

CONDUÇÃO EM AULA

- Peça que a turma escolha uma situação em que começaria pelo logit e outra em que investigaria boosting.
- Discuta hipóteses e as evidências necessárias em cada caso.
- Não revele ainda a tabela final da comparação principal.

CUIDADOS E LIMITES

- Implementações como XGBoost, LightGBM e CatBoost existem e diferem entre si. Aqui não há catálogo de funcionalidades nem recomendação comercial.
- Suporte a dados ausentes e categóricos depende da implementação. Não é propriedade universal do método.
- A frase funciona bem em dados tabulares não substitui um experimento adequado no problema.
- Nenhuma nota subjetiva é atribuída aos algoritmos.

TRANSIÇÃO

Se o modelo usa muitas árvores, como explicar o que ele aprendeu e por que atribuiu determinada PD a um cliente?

---

41 · O que o modelo usa em geral? O que pesou para Bruno?

Experimento sintético; semente e versões no notebook

CONDUÇÃO EM AULA

- Peça a variável mais relevante na carteira e a que mais eleva a previsão de Bruno. Mostre que podem ser diferentes.
- Compare Bruno e Carla mantendo as escalas.
- Termine com a pergunta: se alterarmos essa variável, garantimos essa redução de risco? Resposta: explicação de previsão não identifica efeito causal.

RESPOSTAS

- As duas leituras respondem a perguntas diferentes: uma descreve a carteira, a outra descreve uma previsão.
- A soma da referência com as contribuições reproduz o escore do cliente, o que é conferido numericamente.

CUIDADOS E LIMITES

- Importância por permutação tem unidade própria, aqui queda de AUC na validação. Não é a mesma escala das contribuições locais, que estão em escore.
- Variáveis correlacionadas afetam atribuição e importância.
- Importância por redução de impureza tem limitações distintas das da permutação.
- As contribuições locais são valores de Shapley com expectativa intervencional sobre uma amostra de referência, calculados no notebook, e não um mecanismo econômico.

TRANSIÇÃO

Vamos verificar se conseguimos separar a construção da previsão de sua interpretação.

42 · O novo cliente caiu no grupo B. Qual é a previsão?

Miniatura didática de 10 registros; cálculo próprio

CONDUÇÃO EM AULA

- Dê um minuto individual e outro em duplas.
- Peça que verbalizem onde o desfecho y foi usado e onde deixou de ser necessário.
- Conclua o bloco conectando logit e boosting pela saída probabilística, com funções de escore diferentes.

RESPOSTAS

- Soma em escore: −1,386294 mais 0,200000 mais 0,166078 é igual a −1,020217. A PD é 26,4985%.
- Somar percentuais daria 20% mais 20% mais 16,6078%, o que não é a operação do modelo.
- A média das duas folhas ignoraria o escore inicial e a natureza aditiva da construção.
- Sobre causalidade: uma contribuição positiva do histórico não prova que remover o atraso do cadastro reduziria o risco real.
- Reduzir a taxa de aprendizagem exige recalcular as rodadas e os resíduos, não dividir a PD.

CUIDADOS E LIMITES

- A soma das árvores acontece na escala de log odds. Somar contribuições em probabilidade dá resultado errado.
- A miniatura tem 10 registros e serve para ver o mecanismo. Não é modelo ajustado nem sustenta comparação de desempenho.
- Reconstruir a previsão de um cliente novo não valida o modelo: validação exige amostra fora do tempo, no bloco seguinte.

TRANSIÇÃO

Sabemos como as três técnicas funcionam. Agora precisamos comparar as três em condições iguais e verificar o que cada métrica responde.

---

43 · Estamos comparando modelos ou comparando condições diferentes?

Experimento sintético; semente e versões no notebook

CONDUÇÃO EM AULA

- Recupere o slide 05 em miniatura na tira inferior.
- Pergunte por que usar todos os registros em todos os modelos ainda não garante justiça.
- Revele processamento e seleção, e só depois habilite a abertura dos resultados.

RESPOSTAS

- Boosting com variável futura: a comparação mede vazamento, não algoritmo. Correção: remover a variável e refazer o treinamento.
- Logit sem tratamento enquanto outros receberam engenharia: baseline artificialmente fraco. Correção: dar ao logit transformações e interações plausíveis, como a sensibilidade deste experimento.
- Configuração escolhida pela métrica do teste: o teste deixa de ser independente. Correção: escolher na validação e usar o teste uma única vez.

CUIDADOS E LIMITES

- Modelos especificados antes do teste podem ser comparados nele. Ajustes motivados por esse teste exigem nova avaliação independente.
- A exploração em aula não deve produzir um campeão otimizado depois de ver o resultado.
- Calibradores e cortes pertencem às partições definidas para isso.

TRANSIÇÃO

A primeira pergunta é se o modelo ordena corretamente os clientes de maior e menor risco.

44 · Quem aparece primeiro na fila de risco?

Microexemplo de ranking e experimento sintético, em abas identificadas

CONDUÇÃO EM AULA

- Comece pelo par, não pela integral.
- Pergunte se o escore 0,8 significa 80% de PD calibrada. Neste exemplo não há essa evidência.
- Revele as curvas como resumo da ordenação e diferencie taxa de falsos positivos de inadimplência entre aprovados.

RESPOSTAS

- São 9 pares, 7 corretamente ordenados: AUC igual a 7 dividido por 9, aproximadamente 0,7778.
- Empates contam meio par.
- Mudar o corte não muda a AUC: ela percorre todos os cortes.

CUIDADOS E LIMITES

- Orientação declarada: escore maior significa maior risco.
- AUC por pares não é acurácia, taxa de acerto de aprovados nem rentabilidade.
- KS e AUC resumem aspectos diferentes: um não determina o outro.
- Diferenças pequenas entre modelos podem exigir avaliação de incerteza.

TRANSIÇÃO

Mesmo que o ranking esteja certo, precisamos perguntar se uma previsão de 10% corresponde a cerca de 10% de eventos.

---

45 · Entre os clientes com PD próxima de 10%, quantos ficaram inadimplentes?

Experimento sintético; semente e versões no notebook

CONDUÇÃO EM AULA

- Leia um ponto completo: probabilidade média, proporção observada, n e eventos.
- Aplique a perturbação das odds e pergunte o que deve acontecer com a AUC antes de revelar.
- Discuta por que PD em escala errada distorce perda esperada e política de corte.

RESPOSTAS

- A transformação multiplica as odds por dois e preserva a ordem, portanto a AUC permanece idêntica.
- A calibração e as perdas mudam, e a direção depende de como o modelo já estava calibrado.

CUIDADOS E LIMITES

- Wilson por faixa é descrição binomial aproximada e não incorpora toda a seleção e modelagem.
- Brier e log loss avaliam qualidade probabilística, mas não são medidas exclusivas de calibração.
- Número de faixas, calibrador e versão não devem ser escolhidos pelo teste final.
- O calibrador de Platt foi ajustado na partição de calibração e aplicado ao teste sem reajuste.

APROFUNDAMENTO

- O calibrador de Platt foi ajustado na partição de calibração e aplicado ao teste sem reajuste. O efeito observado na tela é o desse procedimento, não de um recalibrador ajustado no próprio teste.
- Platt é uma logística sobre o escore: $p_calibrada = \sigma(a \cdot z + b)$. Como a transformação é monotônica em z quando a é positivo, ela não altera a ordem nem a AUC, e é exatamente por isso que serve para corrigir nível sem estragar ordenação.

TRANSIÇÃO

Uma PD só vira ação quando definimos uma política. Vamos observar como um corte altera a carteira aprovada.

46 · Até qual PD vamos aprovar?

Experimento sintético; semente e versões no notebook

CONDUÇÃO EM AULA

- Reduza o corte e observe volume e composição.
- Pergunte se menor inadimplência implica melhor resultado. Não: pode haver perda de receita.
- Compare dois modelos no mesmo corte e depois no mesmo volume, e explique que são perguntas diferentes.

RESPOSTAS

- Taxa de aprovação é o número de aprovados dividido pelo total avaliado.
- Inadimplência entre aprovados é a soma de eventos entre aprovados dividida pelo número de aprovados. Contratos recusados não entram no denominador.
- Média das PDs aprovadas e taxa observada são coisas diferentes: a primeira é previsão, a segunda é desfecho.

CUIDADOS E LIMITES

- A avaliação usa contratos com desfecho conhecido da população sintética, o que permite simular recusa. Em dados reais de apenas concedidos, inferir o desempenho dos rejeitados exige hipóteses adicionais.
- Limite de crédito, preço e prazo permanecem fixos aqui, para isolar a decisão de aprovação.
- Empates na PD, comuns na árvore, seguem a regra de aprovar todos no limite. Nenhum desempate usa y.
- Se ninguém for aprovado, a inadimplência entre aprovados fica indefinida.

TRANSIÇÃO

Para escolher um corte, precisamos colocar receitas e perdas na mesma conta.

---

47 · Com estas hipóteses, o equilíbrio ocorre em PD de 20%

Economia didática declarada na tela; cálculo próprio

CONDUÇÃO EM AULA

- Construa a perda esperada, depois o resultado, depois o equilíbrio.
- Pergunte o efeito de aumentar a fração perdida mantendo a PD.
- Mostre que o mesmo cliente pode ser aceitável sob uma estrutura econômica e inviável sob outra, sem defender simplesmente aumento de taxa.

RESPOSTAS

- Perda esperada é PD vezes LGD vezes EAD. Em PD de 10%: 0,10 vezes 60% vezes R$ 10.000, igual a R$ 600.
- Resultado esperado é margem menos perda esperada: R$ 1.200 menos R$ 600, igual a R$ 600.
- Equilíbrio: margem dividida por LGD vezes EAD, igual a 20%.

CUIDADOS E LIMITES

- A margem é considerada constante e já líquida dos custos definidos, apenas nesta simplificação. Não é receita recebida em todos os cenários do mundo real.
- Perda esperada não é lucro, e EAD não é o valor original do contrato sem a hipótese de igualdade.
- A versão alternativa, com margem condicional ao adimplemento, tem outro corte e não deve ser misturada a este gráfico.
- Resultados realizados usam y e são diferentes do valor esperado calculado com p.

TRANSIÇÃO

Antes de implantar, ainda precisamos saber se a relação continua válida no tempo e se conseguimos operar essa relação com controle.

48 · O que conseguimos monitorar agora? O que exige esperar?

Experimento sintético; semente e versões no notebook

CONDUÇÃO EM AULA

- Pergunte se PD média crescente prova piora do modelo. Pode ser mudança da população.
- Pergunte se ausência de inadimplência recente prova segurança. Pode ser falta de maturação.
- Relacione à definição do alvo do slide 03 e termine com a necessidade de versionar modelo, dados, regra e explicações.

RESPOSTAS

- Imediatos: distribuição de renda e comprometimento, ausentes, taxa de aprovação e PD média.
- Com alvo maturado: inadimplência observada, calibração, discriminação e resultado.
- A data de observação define o que pode ser medido: uma coorte só entra nas métricas de desfecho 12 meses depois da contratação.

CUIDADOS E LIMITES

- PSI, se citado, é indicador de mudança de distribuição e depende de faixas e tamanho de amostra. Não use limiares universais como regra automática.
- Mudança de distribuição não demonstra por si só perda de desempenho.
- Os cenários são sintéticos e construídos para a aula. Não são previsão macroeconômica nem backtest real.
- Nem todo alerta pede retreino: a ação depende do diagnóstico.

TRANSIÇÃO

Temos desempenho, probabilidades, economia e condições de operação. Agora vamos tomar uma decisão de comitê.

---

49 · Defenda sua recomendação para o comitê

Experimento sintético; semente e versões no notebook

CONDUÇÃO EM AULA

- Divida a turma em três grupos ou peça decisões individuais.
- Cada grupo tem dois minutos para recomendar e um minuto para defender.
- Solicite uma razão que faria mudar de opinião.
- Reforce que uma evidência de teste não autoriza ajustar repetidamente o modelo no mesmo teste e chamar isso de avaliação independente.

CUIDADOS E LIMITES

- Os números são de base sintética e servem à aprendizagem, não a uma decisão real de concessão.
- Valores próximos não autorizam declarar superioridade conclusiva.
- O formulário fica apenas neste dispositivo e não envia nada.
- Nenhum custo monetário de implantação é atribuído aos modelos: eles não foram medidos.

TRANSIÇÃO

Independentemente da escolha, quais princípios devem acompanhar qualquer modelagem de crédito?

50 · Você consegue explicar, comparar e usar os três modelos?

Síntese do curso

CONDUÇÃO EM AULA

- Peça que um aluno resuma cada técnica em uma frase.
- Retome a decisão do slide 01 sem inventar quais respostas a turma deu. Se as escolhas locais foram salvas, identifique que são escolhas naquele dispositivo.
- Pergunte o que agora exigiriam antes de aprovar uma mudança de modelo.

RESPOSTAS

- Odds vezes 2 não é PD vezes 2: odds e probabilidade são escalas diferentes.
- Sim, a mesma folha agrupa atributos distintos que satisfazem as mesmas regras.
- Não: o boosting logístico soma no escore F e aplica a transformação depois.

CUIDADOS E LIMITES

- Complexidade pode ser útil, e a qualidade da decisão depende do processo inteiro.
- Referências, notebook e modo estudo ficam disponíveis nos recursos finais.
- Não existe um 51º slide de agradecimento: o fechamento é esta síntese.

TRANSIÇÃO

Fim da sequência principal.