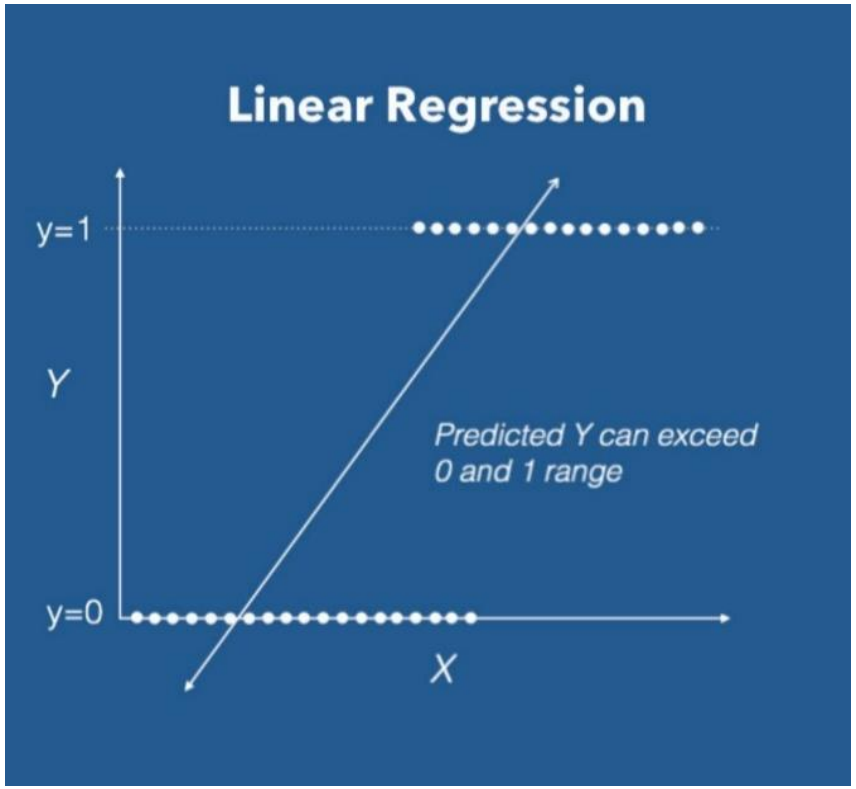
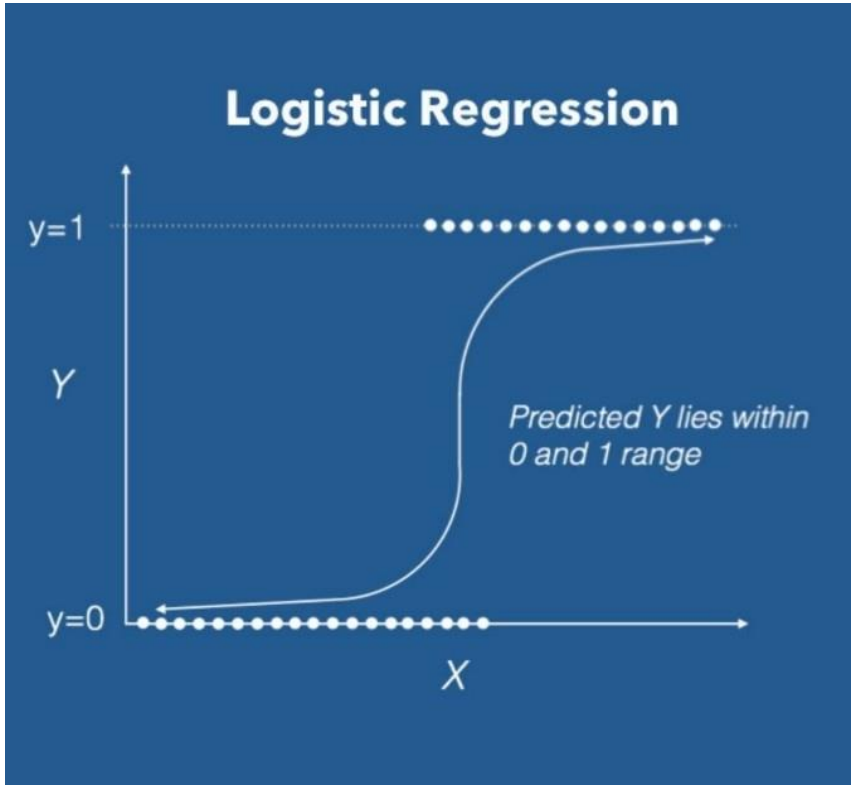
Introdução ao Aprendizado de Máquina:

Modelos de Regressão Linear e Logística

Thales A. P. West, Ph.D.

Algoritmos comuns de aprendizado de máquina





***Regressões lineares** preveem valores numéricos, com base em uma relação linear entre diferentes valores/variáveis*

***Regressões** logísticas fazem previsões para variáveis de resposta categóricas (por exemplo, dados sim/não ou dados 0/1)*

Agenda

Parte 1. Modelos de regressão linear

Parte 2. Modelos de regressão logística

Parte 3. Modelos de regressão não-linear

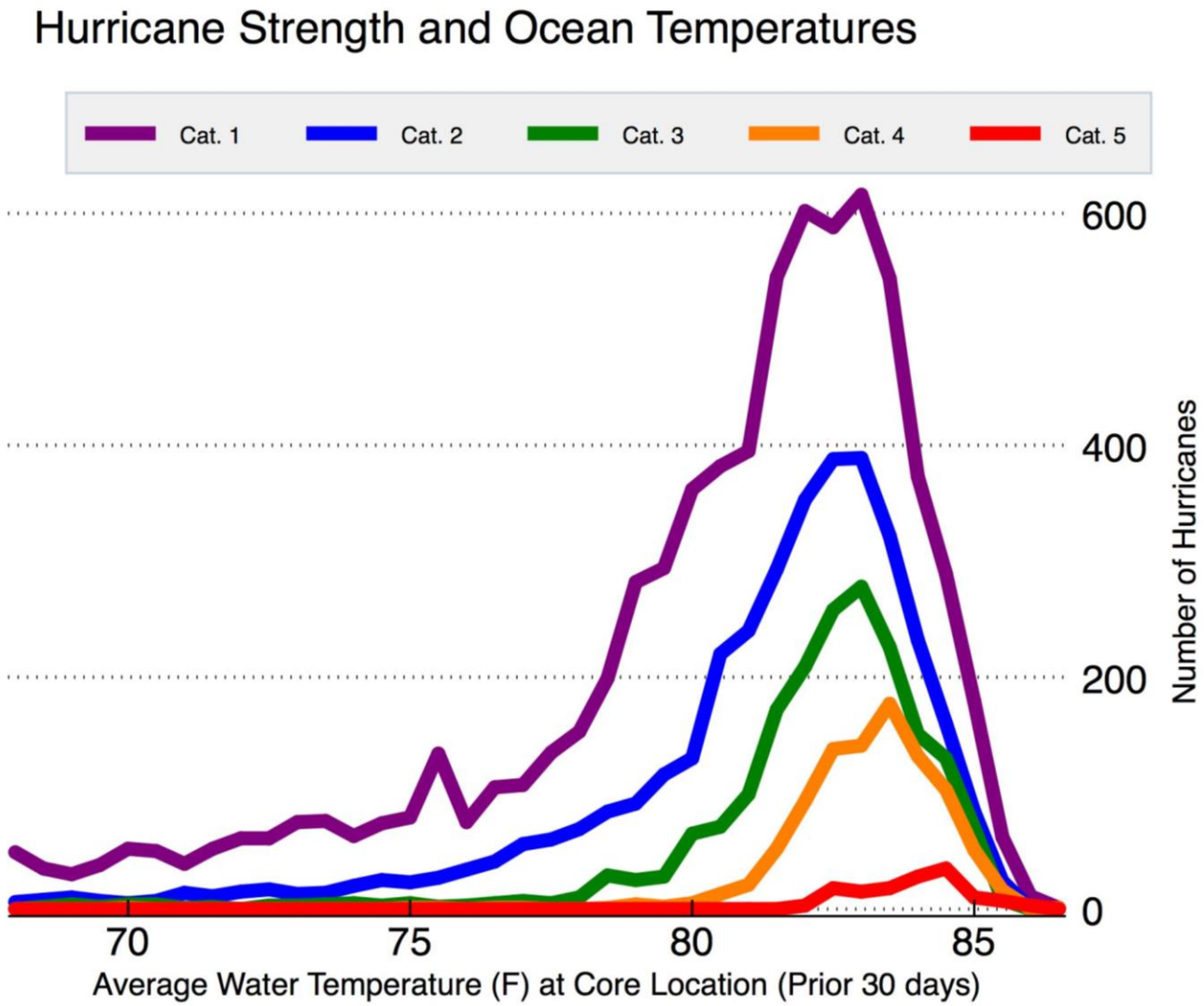
Agenda

Parte 1. Modelos de regressão linear

Parte 2. Modelos de regressão logística

Parte 3. Modelos de regressão não-linear

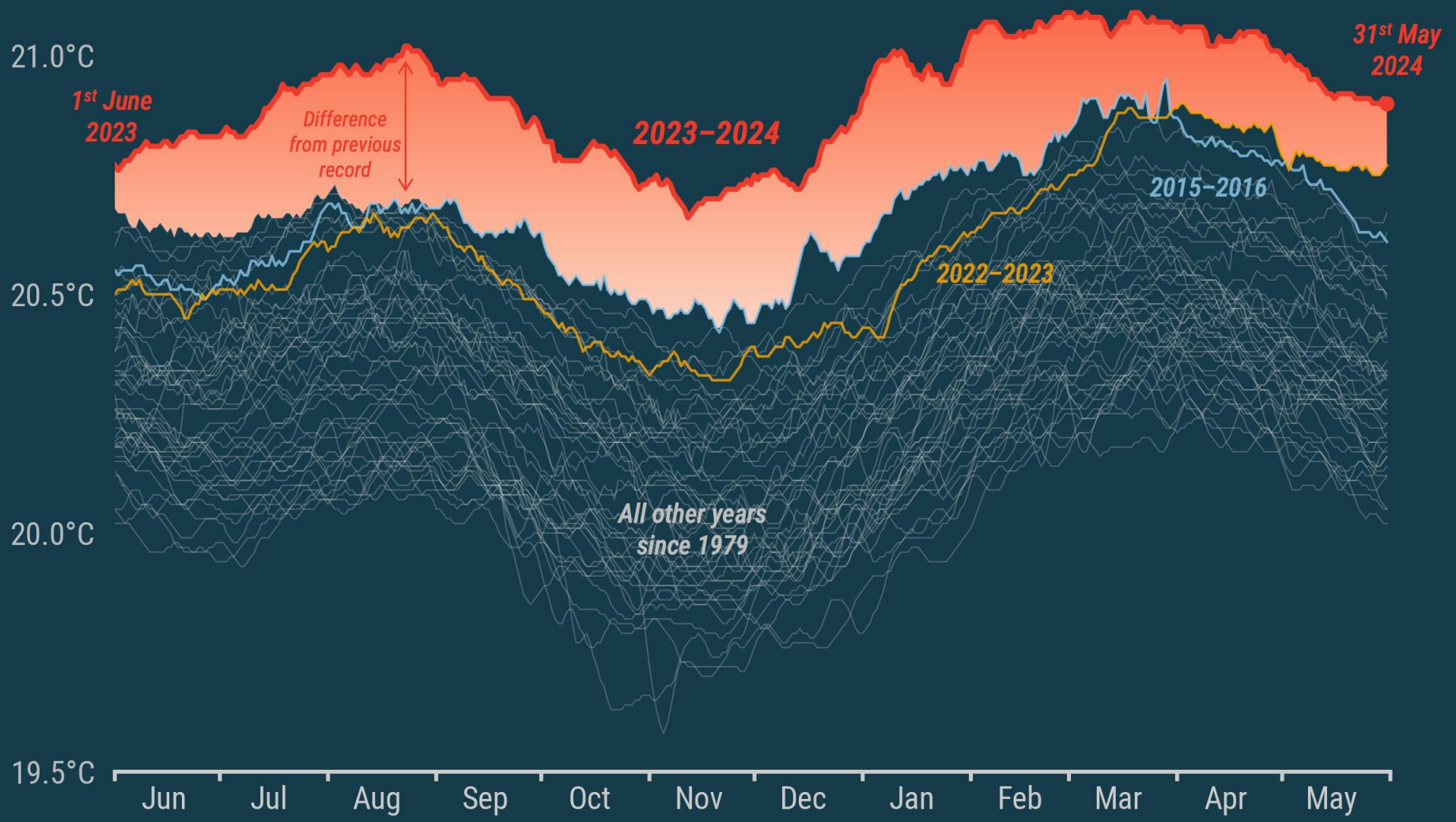
O que fazemos com o aprendizado de máquina?



Kernal density functions of SSTs by hurricane category. Area under each curve represents 100% of hurricanes of that type. Hurricane wind speeds via HURDAT.

Daily sea surface temperature for 60°S-60°N

Data: ERA5 1979–2024 • Credit: C3S/ECMWF



PROGRAMME OF THE EUROPEAN UNION

Copernicus Europe's eyes on Earth

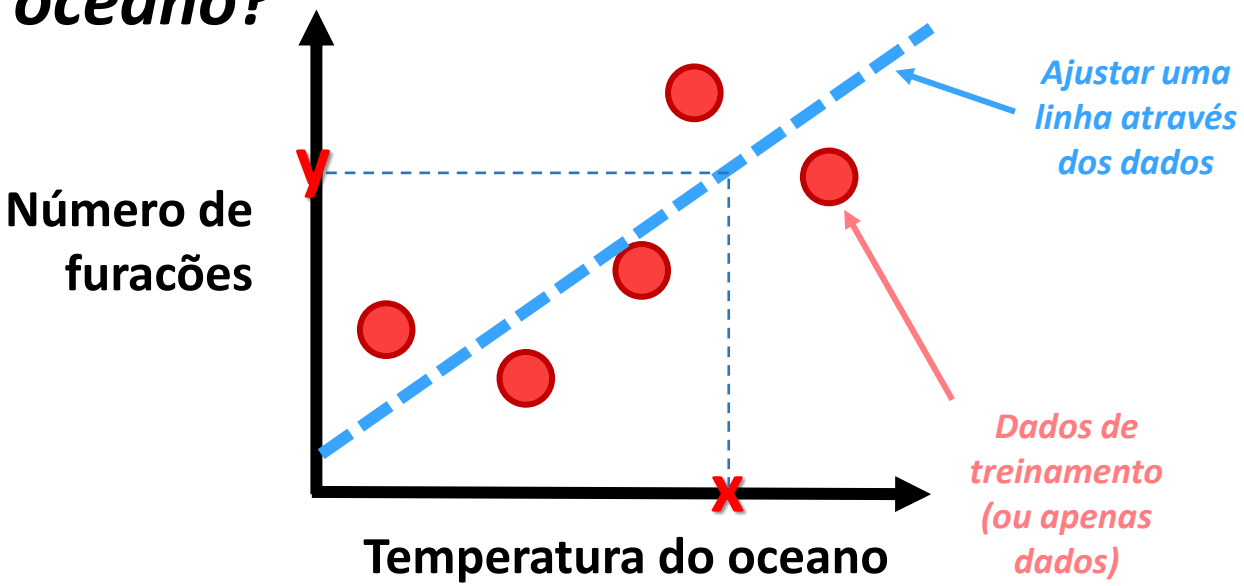
IMPLEMENTED BY ECMWF

Climate Change Service

climate.copernicus.eu

Ajustando um modelo de regressão linear

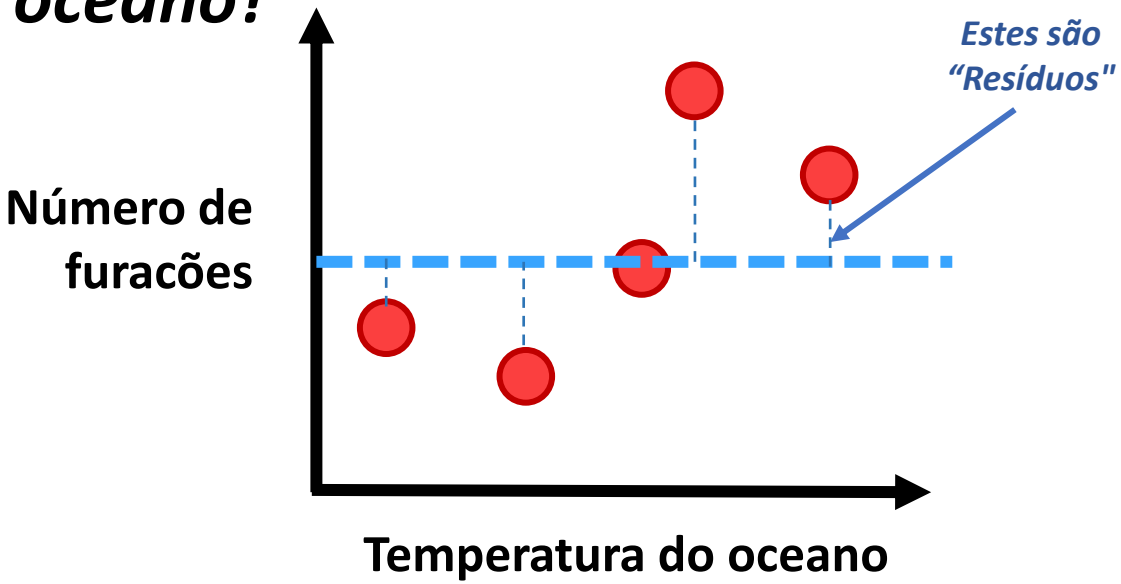
Podemos prever furacões com base na temperatura do oceano?



Podemos usar a linha ajustada para prever o número de furacões (y) com base na temperatura do oceano (x)

Ajustando um modelo de regressão linear

Podemos prever furacões com base na temperatura do oceano?

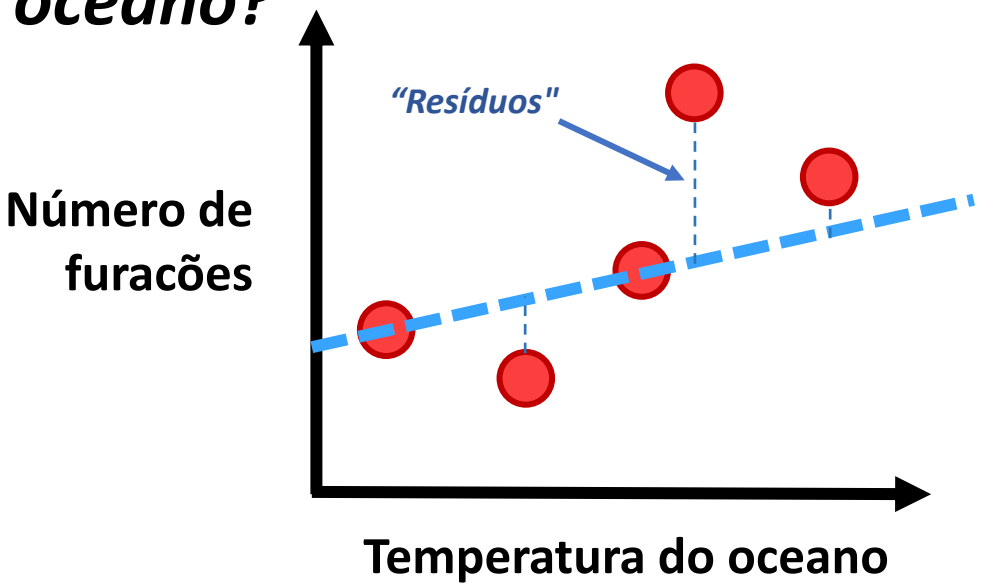


Mas como definir a posição da linha?

Podemos ajustar uma linha através dos dados e medir os resíduos

Ajustando um modelo de regressão linear

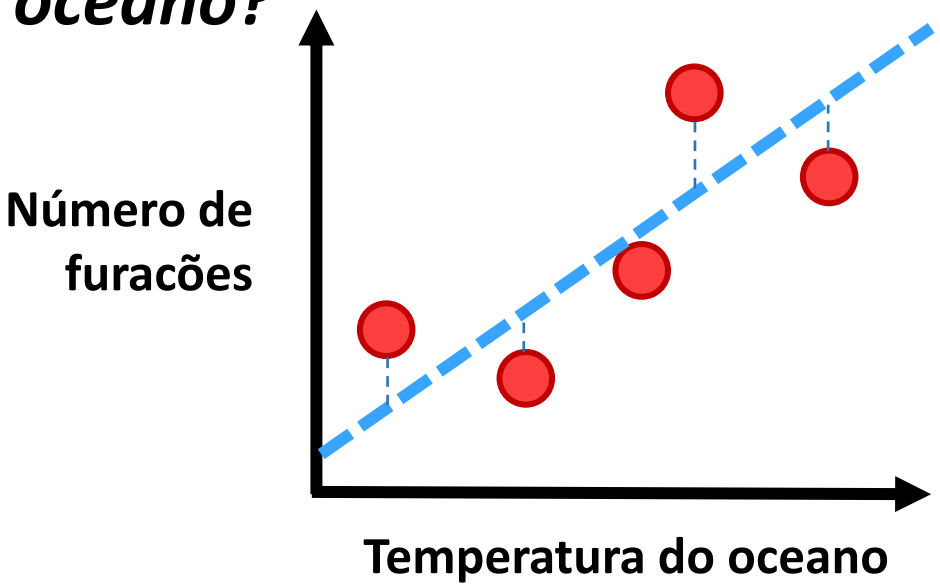
Podemos prever furacões com base na temperatura do oceano?



… E podemos traçar outra linha através dos dados, e fazer isso repetidamente, enquanto continuamos medindo os resíduos a cada vez

Ajustando um modelo de regressão linear

Podemos prever furacões com base na temperatura do oceano?



… Até encontrarmos uma linha com a menor Soma dos Resíduos Quadrados (SRQ)

… mas por que ao quadrado?

Nota: Na Análise de Regressão, a minimização do SRQ é chamada de Método dos Mínimos Quadrados (existem varias formas de minimizar os quadrados!)

Mínimos Quadrados Ordinários (OLS): Tradicional

$$y = \beta_0 + \beta_1 x + \varepsilon \longrightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

→ Solução em "forma fechada" ("closed-form"): fórmula direta (determinística) para calcular os coeficientes

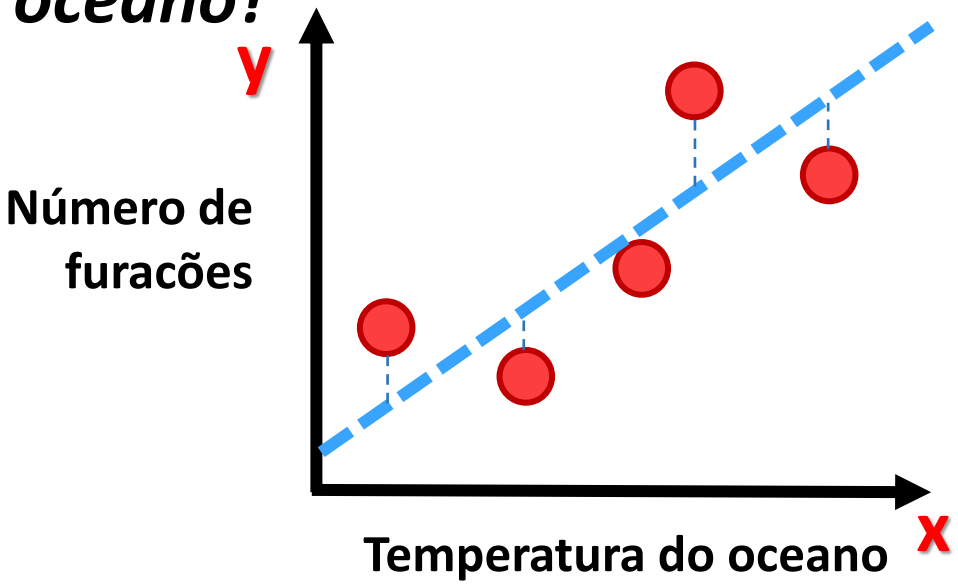
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \longrightarrow \hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Onde:

- $\mathbf{X}$ é a matriz de design (com coluna de 1s para o intercepto),
- $\mathbf{y}$ é o vetor de respostas,
- $\hat{\boldsymbol{\beta}}$ é o vetor de coeficientes estimados.

Ajustando um modelo de regressão linear

Podemos prever furacões com base na temperatura do oceano?



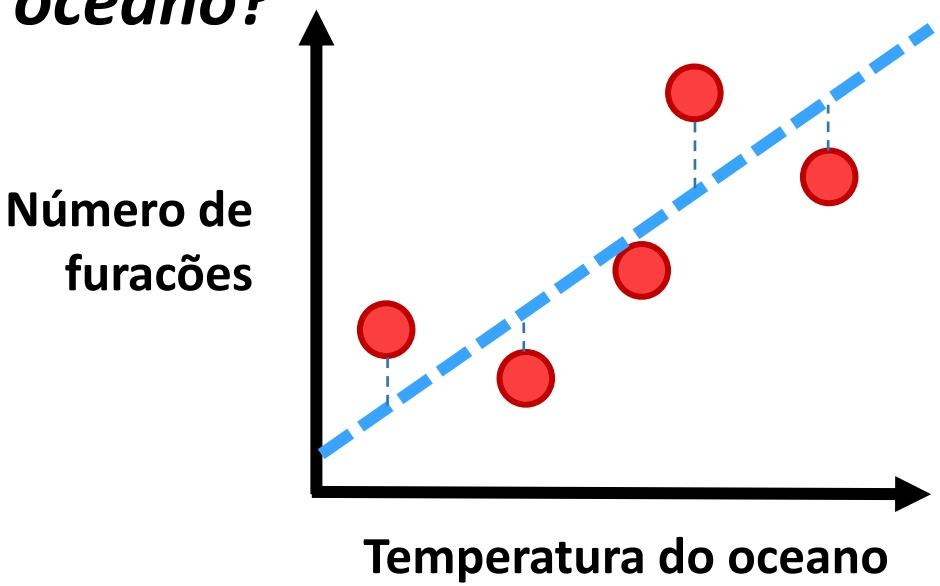
$$y = \underline{0.1} + \underline{0.78} \cdot x$$

interceptação do eixo y

Declive (slope)

Ajustando um modelo linear de forma iterativa

Podemos prever furacões com base na temperatura do oceano?



Para um modelo descrito como:

$$\hat{y}_i = \theta_1 + \theta_2 x_i$$

A minimização entre o valor previsto $(\hat{y}_i)$ ***e o valor observado*** (y_i) ***pode ser descrita como:***

$$minimize \frac{1}{n} = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

Ajustando um modelo linear de forma iterativa

Em "liguagem de Machine Learning", podemos chamar a equação de minimização $minimize\,\frac{1}{n} = \sum_{i=1}^{n}(\hat{y}_i - y_i)^2$ ***de Função de Custo ou Função de Perda***

Os valores de θ_1 ***e*** θ_2 ***do modelo linear*** $\hat{y}_i = \theta_1 + \theta_2 x_i$ ***podem ser calculados iterativamente, geralmente atravez do método Descida de Gradiente ou Gradiente Descendente (Gradient Descent")***

→ ***A idéia é começar a optimização assumindo valores aleatórios para*** θ_1 ***e*** θ_2 ***e atualizalos iterativamente de forma a reduzir a Função de Custo!!!***

Ajustando um modelo linear de forma iterativa

A atualização dos valores de θ_1 e θ_2 é basedo no resultados de derivativas calculadas pela Descida de Gradiente

$$
\begin{aligned}
Eu_{\theta_1} &= \frac{\partial\, J\,(\,\theta_1, \theta_2)}{\partial\, \theta_1} \\
&= \frac{\partial}{\partial\, \theta_1}\left[\frac{1}{n}\left(\sum_{eu=1}^{n}(\hat{e}_{eu} - e_{eu})^2\right)\right] \\
&= \frac{1}{n}\left[\sum_{eu=1}^{n} 2\,(\hat{e}_{eu} - e_{eu})\left(\frac{\partial}{\partial\, \theta_1}(\hat{e}_{eu} - e_{eu})\right)\right] \\
&= \frac{1}{n}\left[\sum_{eu=1}^{n} 2\,(\hat{e}_{eu} - e_{eu})\left(\frac{\partial}{\partial\, \theta_1}(\,\theta_1 + \theta_2 x_{eu} - e_{eu})\right)\right] \\
&= \frac{1}{n}\left[\sum_{eu=1}^{n} 2\,(\hat{e}_{eu} - e_{eu})\,(\,1 + 0 - 0\,)\right] \\
&= \frac{1}{n}\left[\sum_{eu=1}^{n} (\hat{e}_{eu} - e_{eu})\,(\,2\,)\right] \\
&= \frac{2}{n}\sum_{eu=1}^{n}(\hat{e}_{eu} - e_{eu})
\end{aligned}
$$

$$
\begin{aligned}
Eu_{\theta_2} &= \frac{\partial\, J\,(\,\theta_1, \theta_2)}{\partial\, \theta_2} \\
&= \frac{\partial}{\partial\, \theta_2}\left[\frac{1}{n}\left(\sum_{eu=1}^{n}(\hat{e}_{eu} - e_{eu})^2\right)\right] \\
&= \frac{1}{n}\left[\sum_{eu=1}^{n} 2\,(\hat{e}_{eu} - e_{eu})\left(\frac{\partial}{\partial\, \theta_2}(\hat{e}_{eu} - e_{eu})\right)\right] \\
&= \frac{1}{n}\left[\sum_{eu=1}^{n} 2\,(\hat{e}_{eu} - e_{eu})\left(\frac{\partial}{\partial\, \theta_2}(\,\theta_1 + \theta_2 x_{eu} - e_{eu})\right)\right] \\
&= \frac{1}{n}\left[\sum_{eu=1}^{n} 2\,(\hat{e}_{eu} - e_{eu})\,(\,0 + x_{eu} - 0\,)\right] \\
&= \frac{1}{n}\left[\sum_{eu=1}^{n} (\hat{e}_{eu} - e_{eu})\,(\,2\,x_{eu})\right] \\
&= \frac{2}{n}\sum_{eu=1}^{n}(\hat{e}_{eu} - e_{eu})\cdot x_{eu}
\end{aligned}
$$

Ajustando um modelo linear de forma iterativa

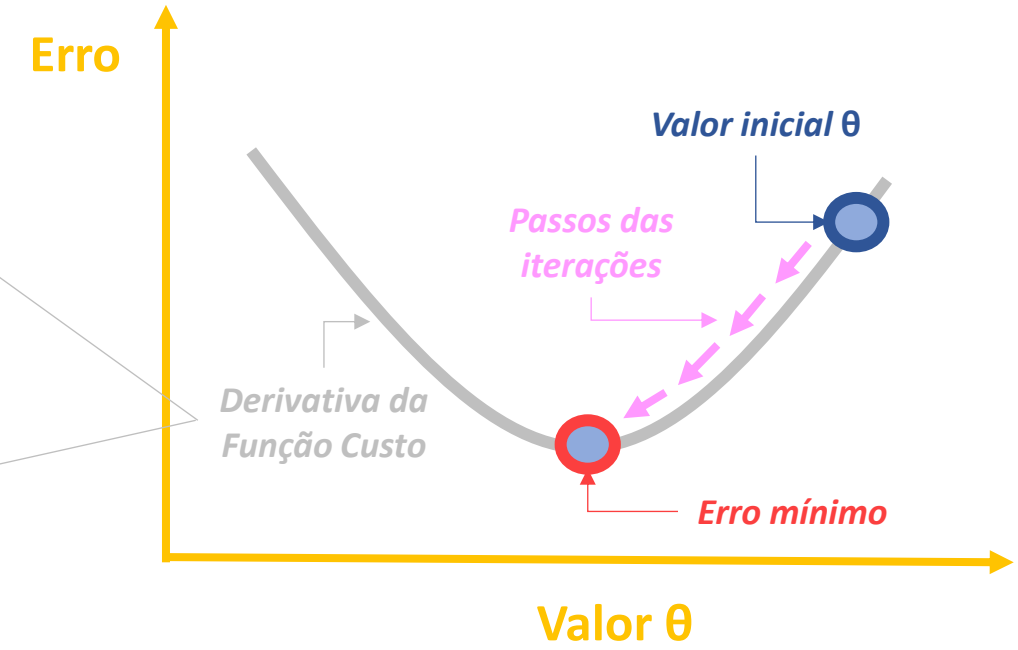
$$\theta_{1_{Atualizado}} = \theta_{1_{Anterior}} - \alpha \frac{\partial\ Função\ de\ Custo}{\partial\ \theta_1}$$

$$= \theta_{1_{Anterior}} - \alpha \left(\frac{2}{n} \sum (\hat{y}_i - y_i) \right)$$

$$\theta_{2_{Atualizado}} = \theta_{2_{Anterior}} - \alpha \frac{\partial Função\ de\ Custo}{\partial\ \theta_2}$$

$$= \theta_{2_{Anterior}} - \alpha \left(\frac{2}{n} \sum (\hat{y}_i - y_i)\, x_i \right)$$

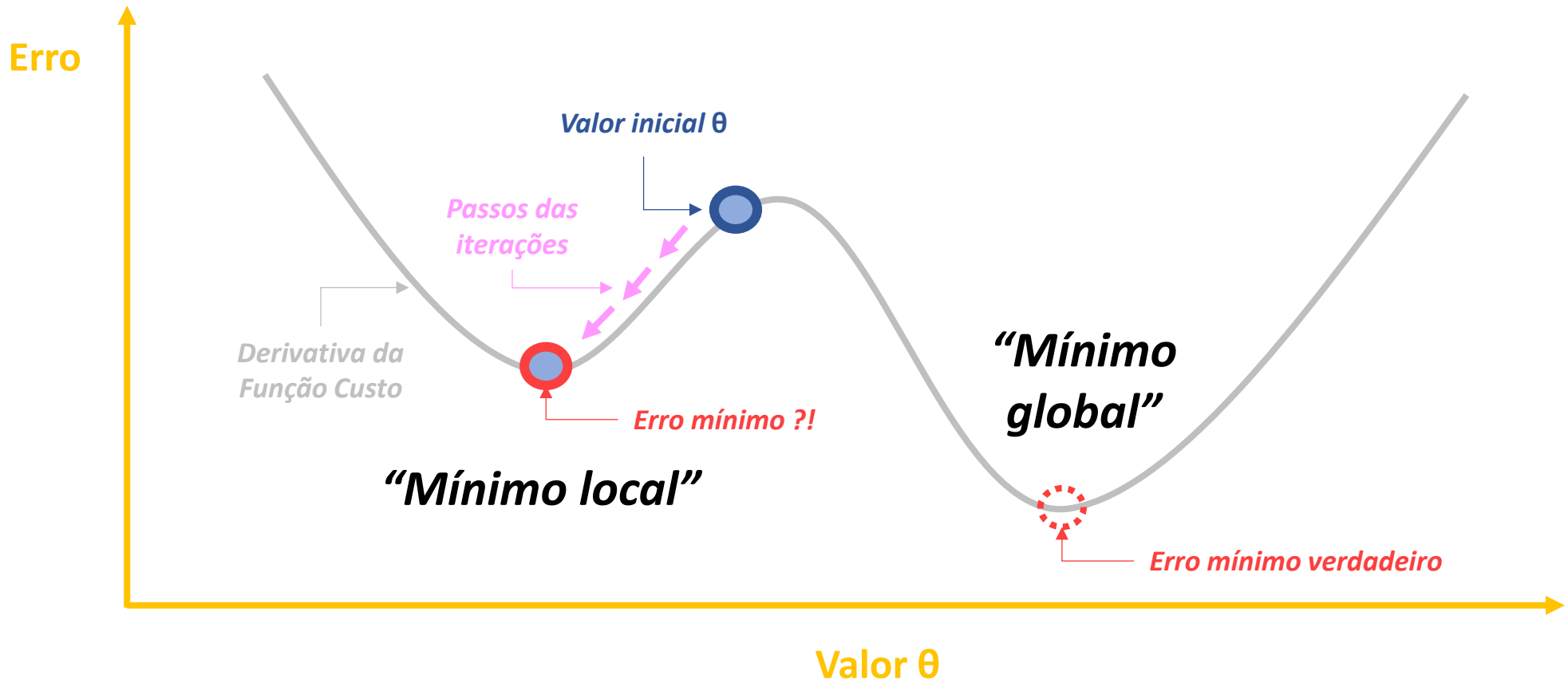
Taxa de aprendizado



Comparação de métodos de ajuste de modelos

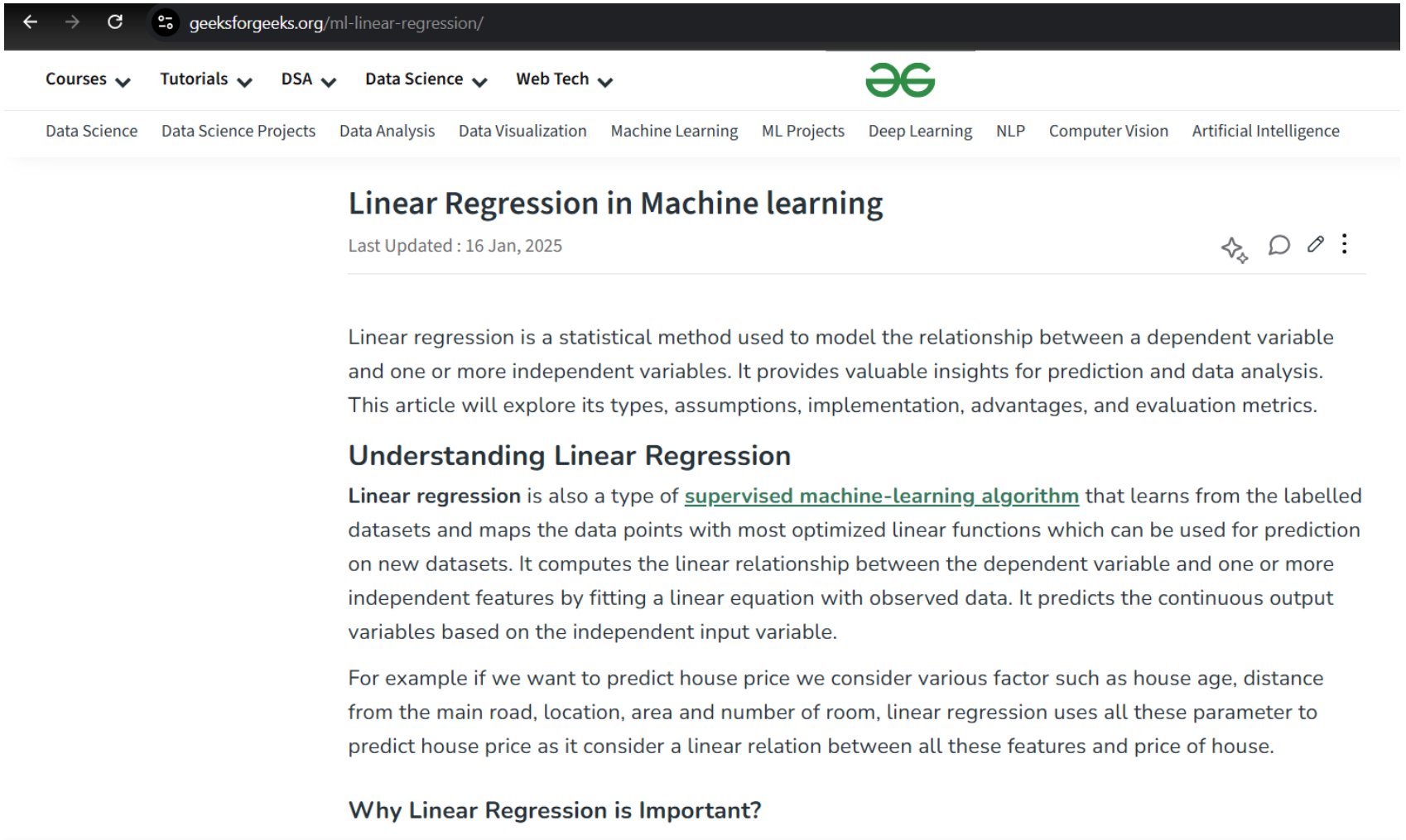
Característica	Mínimos Quadrados Ordinários	Descida de Gradiente
Tipo de Solução	Analítica (determinística)	Iterativa (optimização)
Complexidade Computacional	Pode ser alto para grandes conjuntos de dados	Geralmente mais eficiente para grandes conjuntos de dados
Convergência	Solução exata (em condições ideais)	Converge para uma solução aproximada
Aplicação	Modelos lineares	Modelos lineares e não lineares
Taxa de Aprendizado	N/A	Requer ajuste da taxa de aprendizado
Mínimos Locais	N/A	Pode ficar preso em mínimos locais
Memória	Pode exigir grande quantidade de memória para grandes conjuntos de dados	Geralmente menos exigente em memória
Velocidade	Rápido para conjuntos de dados pequenos a médios	Pode ser mais rápido para grandes conjuntos de dados

Mínimo local & Mínimo global



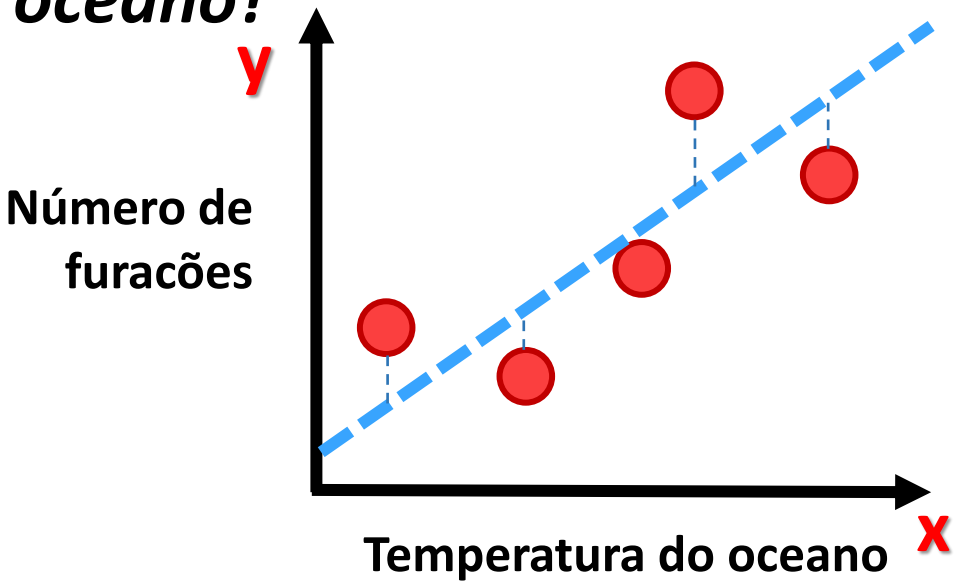
Ajustando um modelo de regressão linear

https://www.geeksforgeeks.org/ml-linear-regression/



Ajustando um modelo de regressão linear

Podemos prever furacões com base na temperatura do oceano?



$$y = \underline{0.1} + \underline{0.78} \cdot x$$

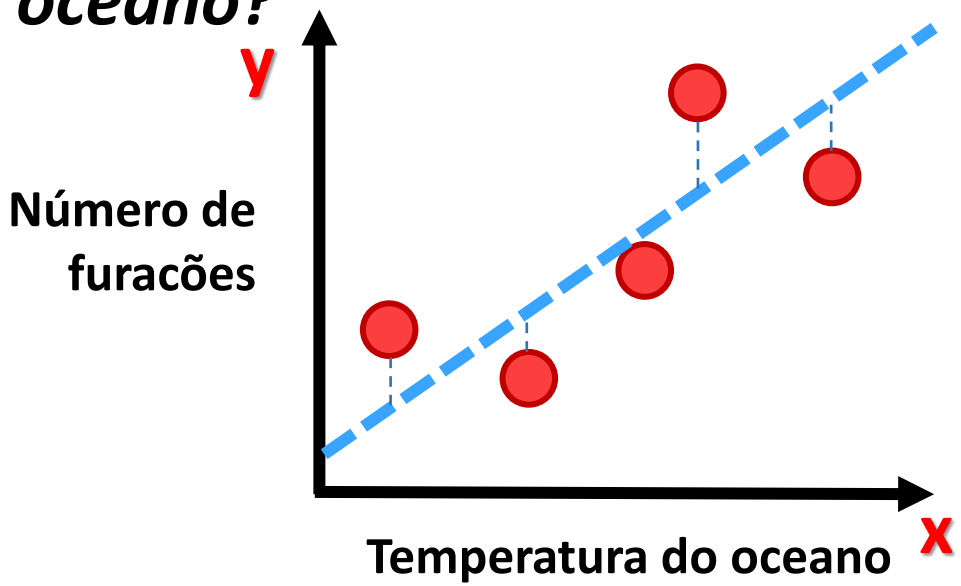
intercepto do eixo y

Declive (slope)

R^2 mede a proporção da variância da variável dependente que é explicada pelas variáveis independentes no modelo

Ajustando um modelo de regressão linear

Podemos prever furacões com base na temperatura do oceano?



R^2 varia de 0 a 1

Por exemplo, um R^2 de 0,6 significaria que 60% da variação no número de furacões pode ser explicada pela temperatura do oceano

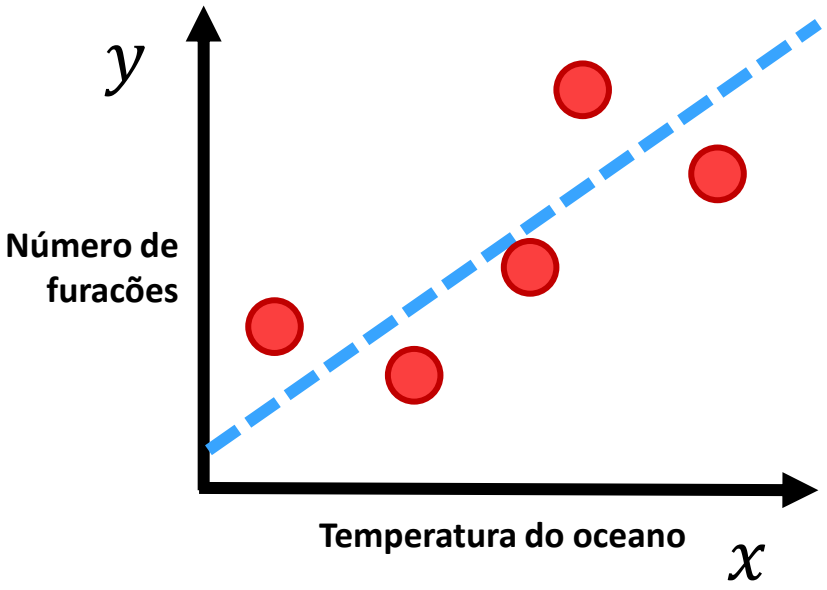
Além do R^2 ...

Uma limitação do R^2 é que ele sempre aumenta quando novas variáveis independentes são adicionadas ao modelo, mesmo que essas variáveis não melhorem significativamente o ajuste*. Isso pode levar a um sobreajuste ("overfitting") e perda de interpretabilidade*

➔ ***O R^2 ajustado corrige a limitação do R^2 ao penalizar a adição de variáveis independentes que não contribuem significativamente para o modelo.*** *Ele leva em consideração o número de variáveis independentes e o tamanho da amostra. O R^2 ajustado fornece uma estimativa mais realista do ajuste do modelo, especialmente quando se trabalha com modelos de regressão múltipla*

➔ ***Critério de Informação de Akaike (AIC) & Critério de Informação Bayesiano (BIC): são métricas que avaliam o equilíbrio entre o ajuste do modelo e a complexidade do modelo.*** *AIC e BIC podem ser usados para comparar modelos com diferentes números de variáveis independentes. Valores menores de AIC e BIC indicam modelos melhores*

Notações para modelos de regressão



Dependent/response/ outcome/output/target variable

Independent/explanatory/ input variable, covariate, regressor, or "feature" (in machine learning)

$$y = \alpha + \beta x + \varepsilon$$

Error term (residuals)

Intercept

Parameter

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon$$

Interceptações e parâmetros estimados (e sua significância) podem ser facilmente interpretados e de grande relevância

Resultados de modelos de regressão linear

Table 1
Parameter estimates for the production function.

	Agricultural production (Mg ha^{-1})			
	all data	outliers excluded	all data	outliers excluded
Labor (persons year^{-1})	0.15***	0.17***	0.16***	0.18***
	(0.05)	(0.05)	(0.05)	(0.04)
Agricultural area reported by the household (ha)	0.45***	0.51***		
	(0.06)	(0.05)		
Agricultural area estimated with remote sensing (ha)			0.40***	0.42***
			(0.06)	(0.06)
Constant	0.97***	0.72***	1.07***	0.99***
	(0.21)	(0.20)	(0.24)	(0.22)
Observations	531	524	536	529
R^2	0.12	0.18	0.09	0.11
Adjusted R^2	0.12	0.17	0.09	0.11
Residual Std. Error	1.02	0.93	1.05	0.98
	(d.f. = 528)	(d.f. = 521)	(d.f. = 533)	(d.f. = 526)
F Statistic	37.66***	51.63***	26.87***	33.99***
	(d.f. = 2; 528)	(d.f. = 2; 521)	(d.f. = 2; 533)	(d.f. = 2; 526)

Note: *p < 0.10; **p < 0.05; ***p < 0.01; standard errors are noted in parentheses; d.f. = degrees of freedom.

Lagged, Autoregressive & Quadratic parameters

$$y_t = \alpha + \beta_1 x_t + \beta_2 \boldsymbol{x_{t-1}} + \varepsilon$$

Lagged variable

$$y_t = \alpha + \beta \boldsymbol{y_{t-1}} + \varepsilon$$

Autoregressive variable

$t\ (time)$	x_t	x_{t-1}
1	10	-
2	7	10
3	15	7
4	13	15
5	8	13
6	4	8

$$y = \alpha + \beta_1 x + \beta_2 \boldsymbol{x^2} + \varepsilon$$

Quadratic variable

$t\ (time)$	y_t	y_{t-1}
1	11	-
2	27	11
3	5	27
4	3	5
5	18	3
6	14	18

Modelos “Lin-Lin”, “Log-Lin” & “Log-Log”

Modelo Lin-Lin (Linear-Linear):

$$y = \alpha + \beta x + \varepsilon$$

Modelo Lin-Log (Linear-Logarítmico):

$$y = \alpha + \beta \cdot \log(x) + \varepsilon$$

Modelo Log-Lin (logarítmico-linear):

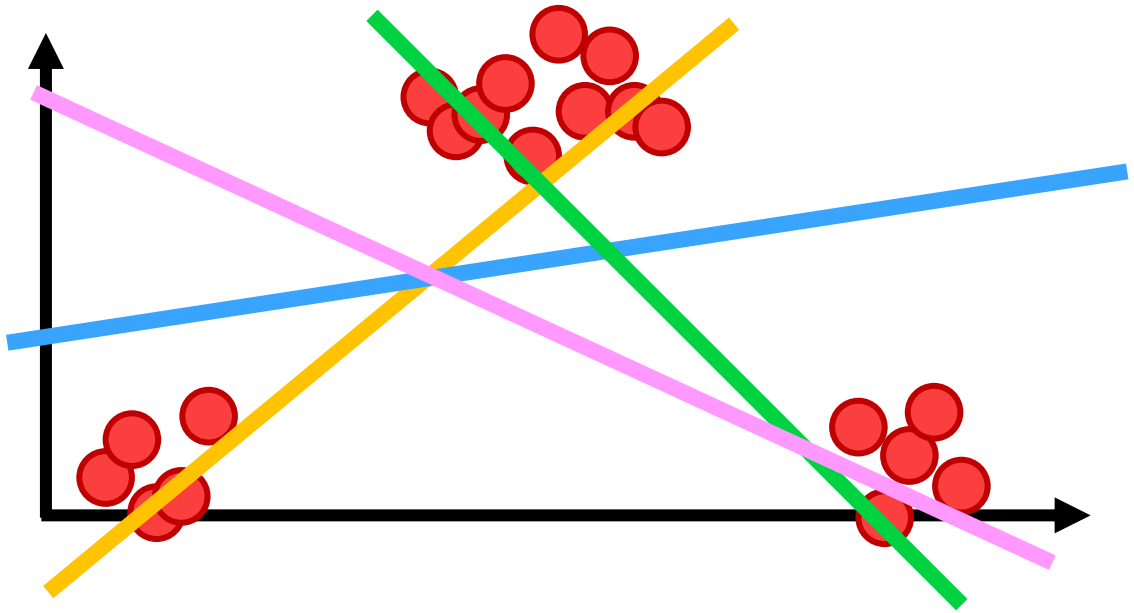
$$\log(y) = \alpha + \beta x + \varepsilon$$

Modelo Log-Log (logarítmico-logarítmico):

$$\log(y) = \alpha + \beta \cdot \log(x) + \varepsilon$$

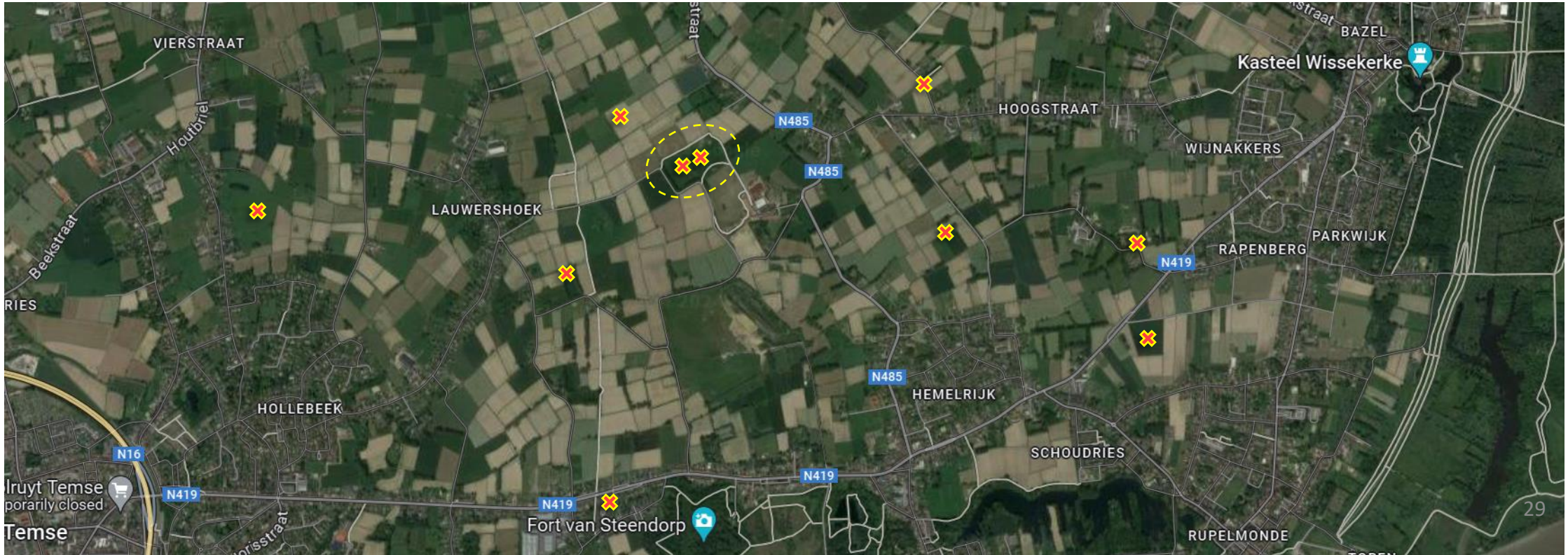
Suposições do modelo de regressão linear

1.Linearidade: A relação entre as variáveis independentes e a variável dependente é considerada linear. Isso significa que as mudanças nas variáveis independentes estão associadas a mudanças constantes na variável dependente



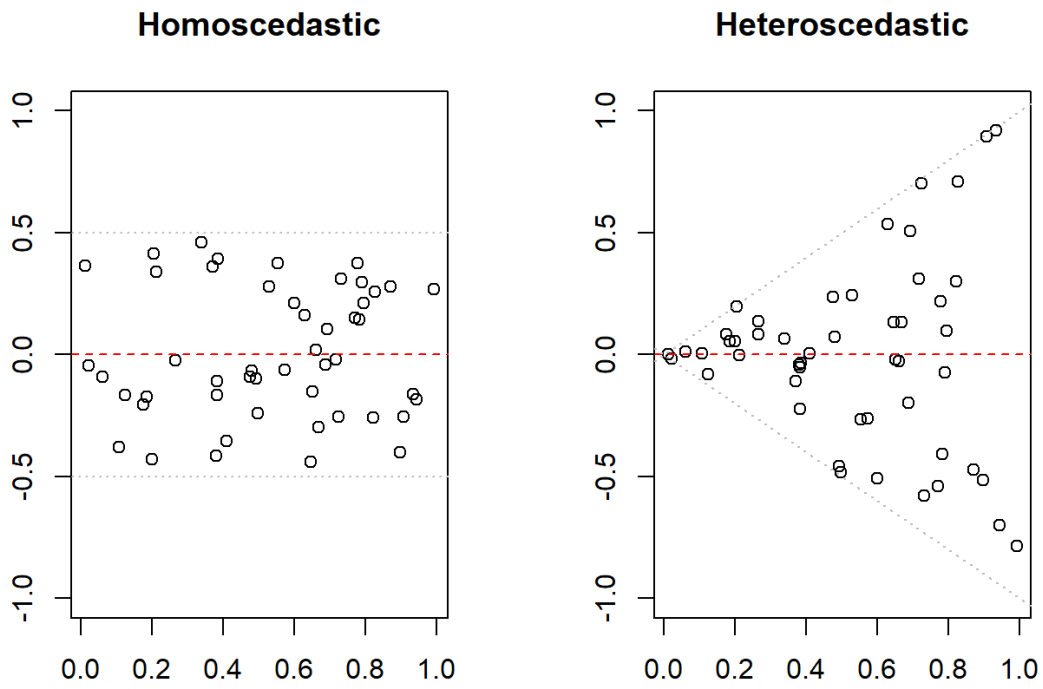
Suposições do modelo de regressão linear

2. Independência: As observações são consideradas independentes umas das outras. Em outras palavras, o valor da variável dependente para uma observação não deve ser influenciado pelos valores de outras observações. A independência é crucial para a significância estatística dos coeficientes



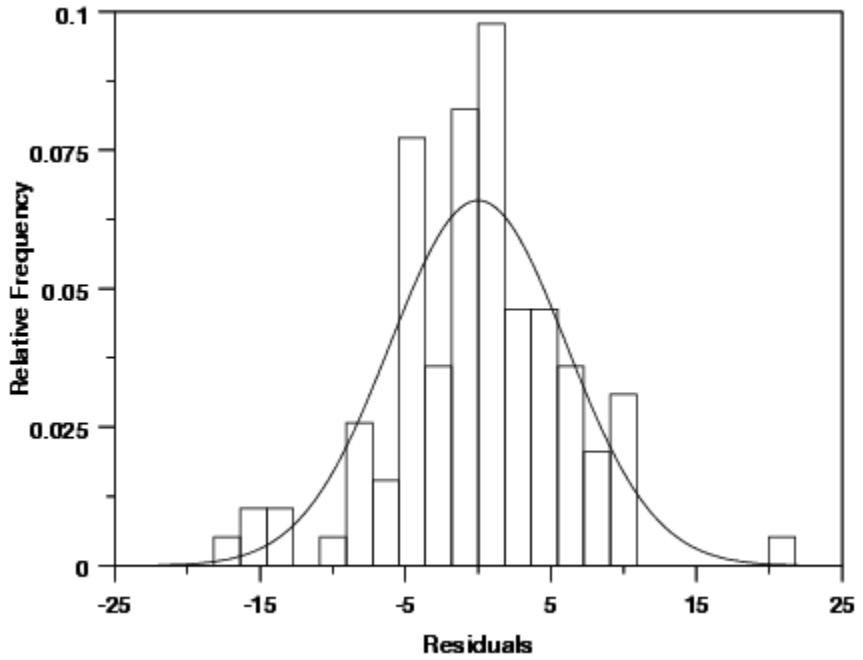
Suposições do modelo de regressão linear

3. Homoscedasticidade: A variância dos resíduos (as diferenças entre os valores observados e previstos) deve ser constante em todos os níveis das variáveis independentes. Em outras palavras, a dispersão dos resíduos deve permanecer aproximadamente a mesma em toda a faixa de valores previstos



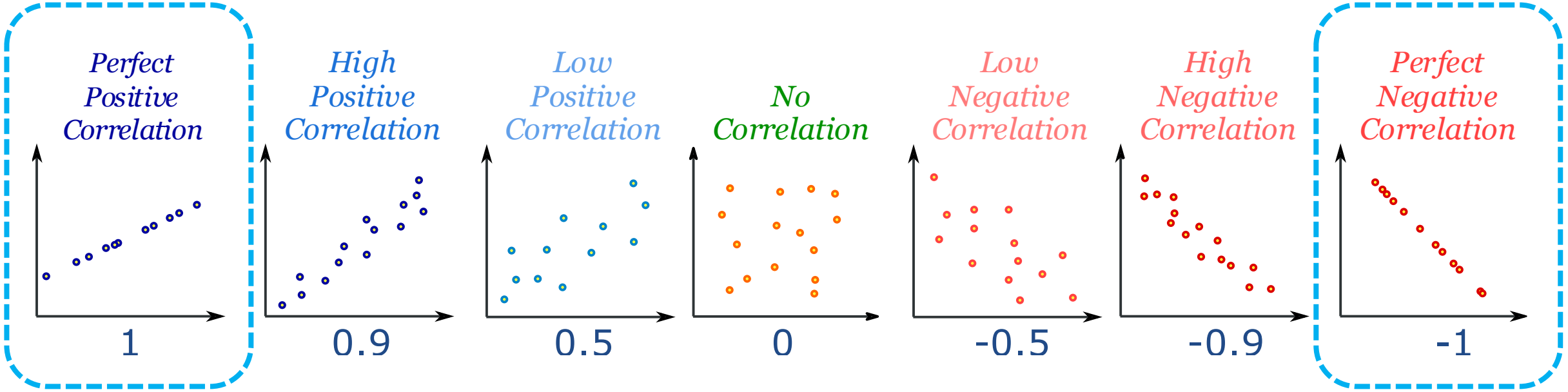
Suposições do modelo de regressão linear

4. Normalidade dos resíduos: Os resíduos (as diferenças entre os valores observados e previstos) devem ser normalmente distribuídos. No entanto, essa suposição não é tão crítica para grandes tamanhos de amostra devido ao Teorema do Limite Central. Para tamanhos de amostra menores, os desvios da normalidade podem afetar a precisão dos intervalos de confiança e dos testes de hipóteses



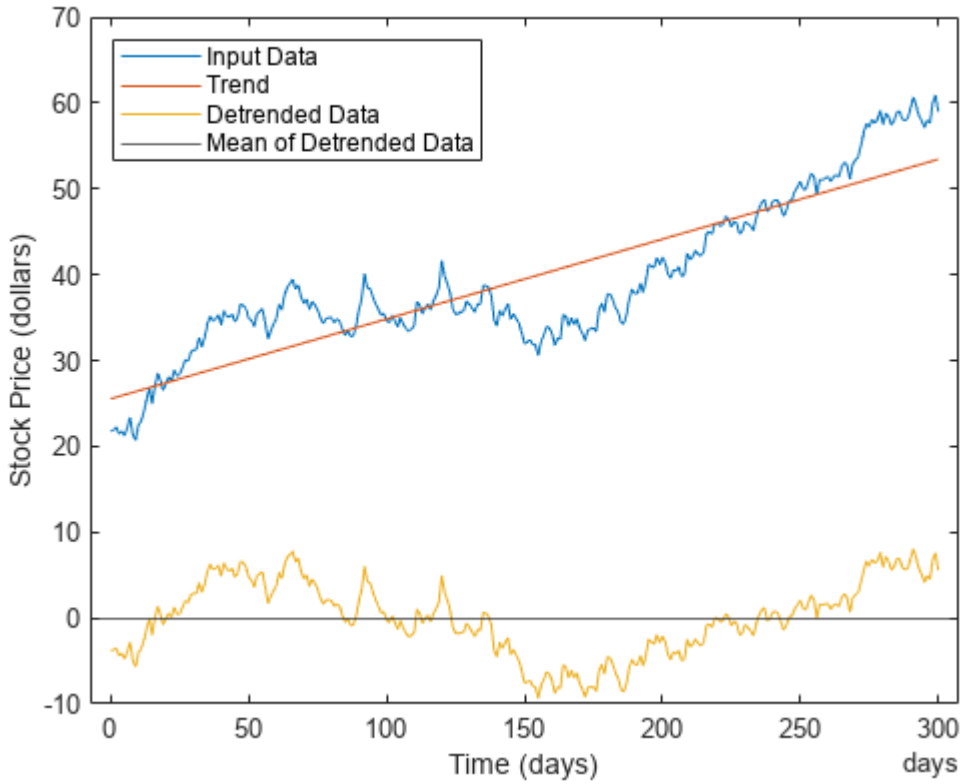
Suposições do modelo de regressão linear

5. Ausência de multicolinearidade: As variáveis independentes não devem estar perfeitamente ou altamente correlacionadas entre si. A multicolinearidade, onde há um alto grau de correlação entre variáveis independentes, pode dificultar o isolamento dos efeitos individuais de cada variável (devido à **redundância de informação** entre elas) → calcular *Fator de Inflação da Variância*



Suposições do modelo de regressão linear

6. Ausência de autocorrelação: Os resíduos não devem exibir padrões ou correlações ao longo do tempo ou no espaço. A autocorrelação, ou a correlação de uma variável consigo mesma em diferentes momentos/locais, pode sugerir que variáveis importantes estão ausentes do modelo



A autocorrelação de resíduos ao longo do tempo é uma das razões pelas quais a redução da tendência ou garantia da estacionariedade é importante na análise de séries temporais. Autocorrelação em resíduos significa que há um padrão ou estrutura restante nos resíduos após o ajuste de um modelo aos dados da série temporal. Esse padrão pode indicar que o modelo não capturou toda a dinâmica subjacente da série temporal. Em outras palavras, quando você tem autocorrelação em resíduos, isso implica que existem relações sistemáticas entre as observações em diferentes pontos de tempo que o modelo não está contabilizando

Mais suposições do modelo de regressão linear...

Aditividade: O modelo pressupõe que o efeito das alterações em uma variável preditora na variável de resposta é consistente, independentemente dos valores das outras variáveis. Essa suposição implica que não há interação entre as variáveis em seus efeitos sobre a variável dependente

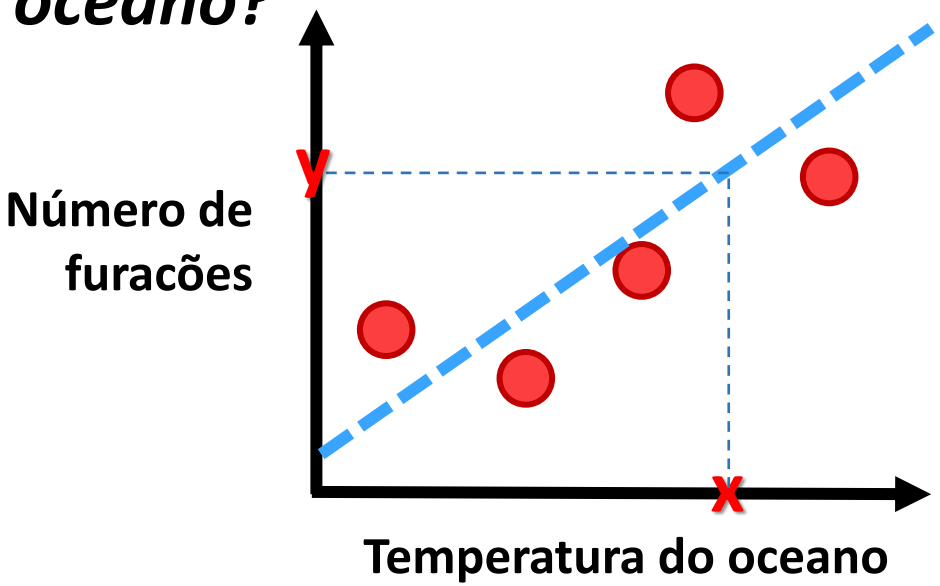
Seleção de recursos ("features"): Na regressão linear múltipla, é essencial selecionar cuidadosamente as variáveis independentes que serão incluídas no modelo. A inclusão de variáveis irrelevantes ou redundantes pode levar ao sobreajuste e complicar a interpretação do modelo

Sobreajuste ("overfitting"): O overfitting ocorre quando o modelo ajusta os dados de treinamento "muito de perto", capturando ruído ou flutuações aleatórias que não representam a verdadeira relação subjacente entre as variáveis. Isso pode levar a um desempenho de generalização ruim em dados novos e não vistos

… mas podemos violar essas suposições?

O aprendizado de máquina tem tudo a ver com fazer previsões e classificações

Podemos prever furacões com base na temperatura do oceano?



Ainda podemos usar um modelo de regressão linear "mal" ajustado para prever o número de furacões (y) com base na temperatura do oceano (x), mas provavelmente nossas previsões não serão muito confiáveis...

Agenda

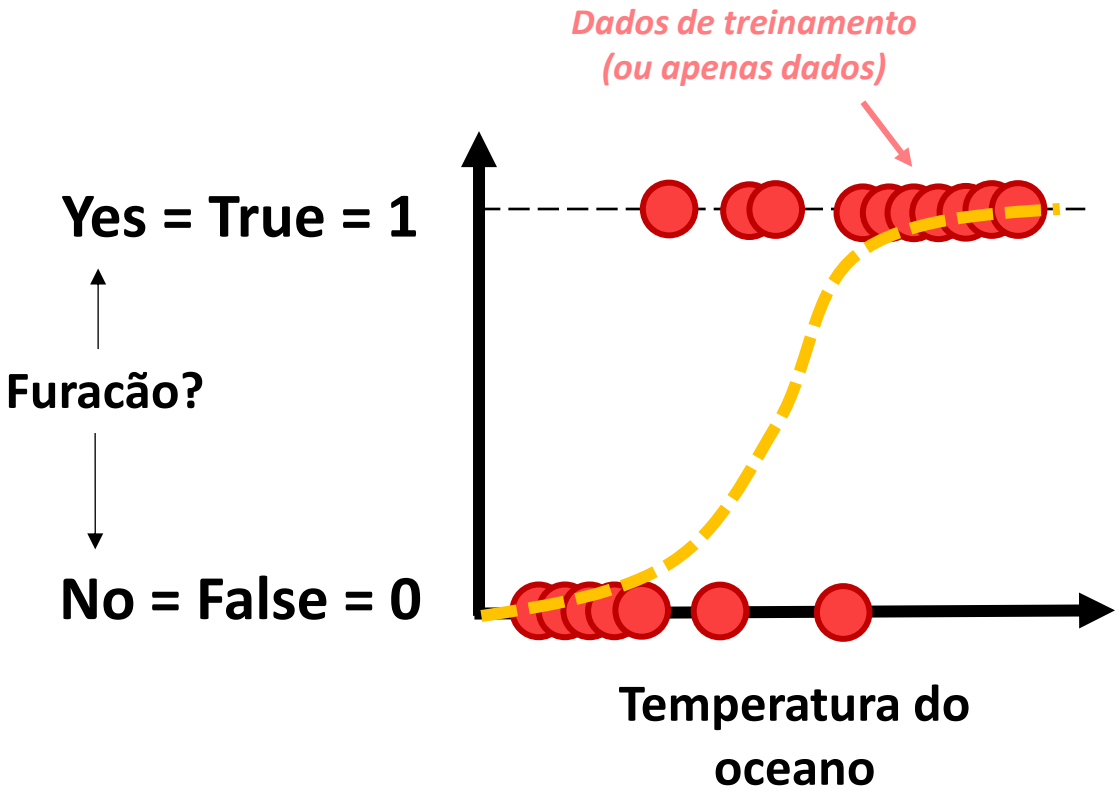
Parte 1. Modelos de regressão linear

Parte 2. Modelos de regressão logística

Parte 3. Modelos de regressão não-linear

Ajustando um modelo de regressão logística

Podemos prever furacões com base na temperatura do oceano?

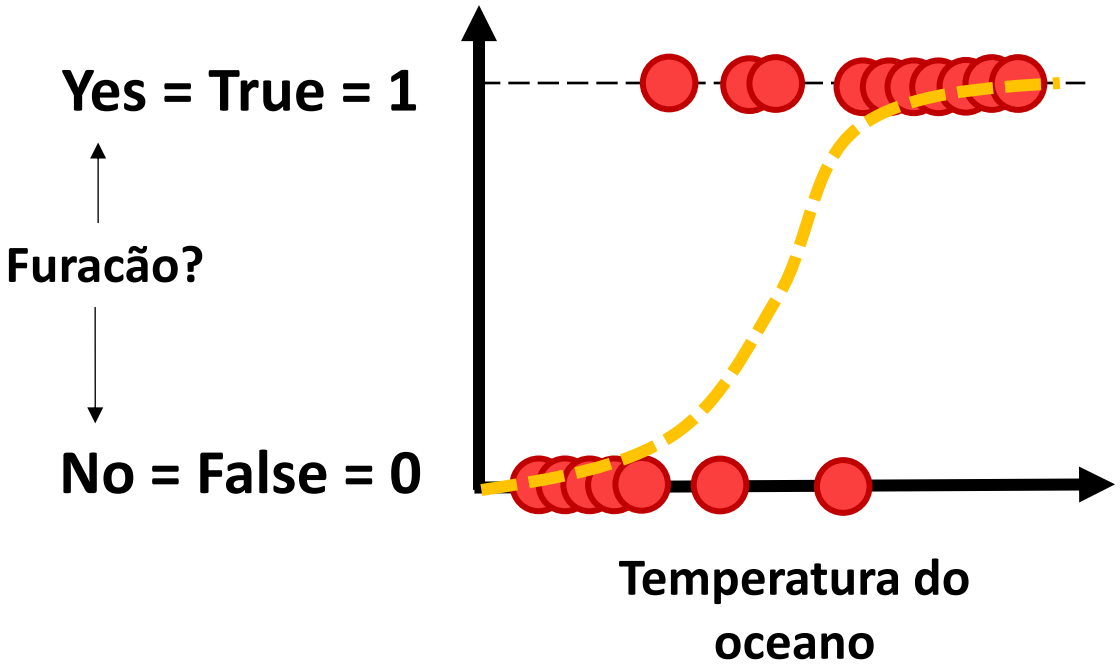


As regressões logísticas funcionam com valores booleanos (verdadeiro/falso, sim/não, 1s/0s, etc.) em vez de valores contínuos

Em vez de ajustar uma linha reta, ela ajusta uma curva "em forma de S" (ou sigmóide) através dos dados

Ajustando um modelo de regressão logística

Podemos prever furacões com base na temperatura do oceano?

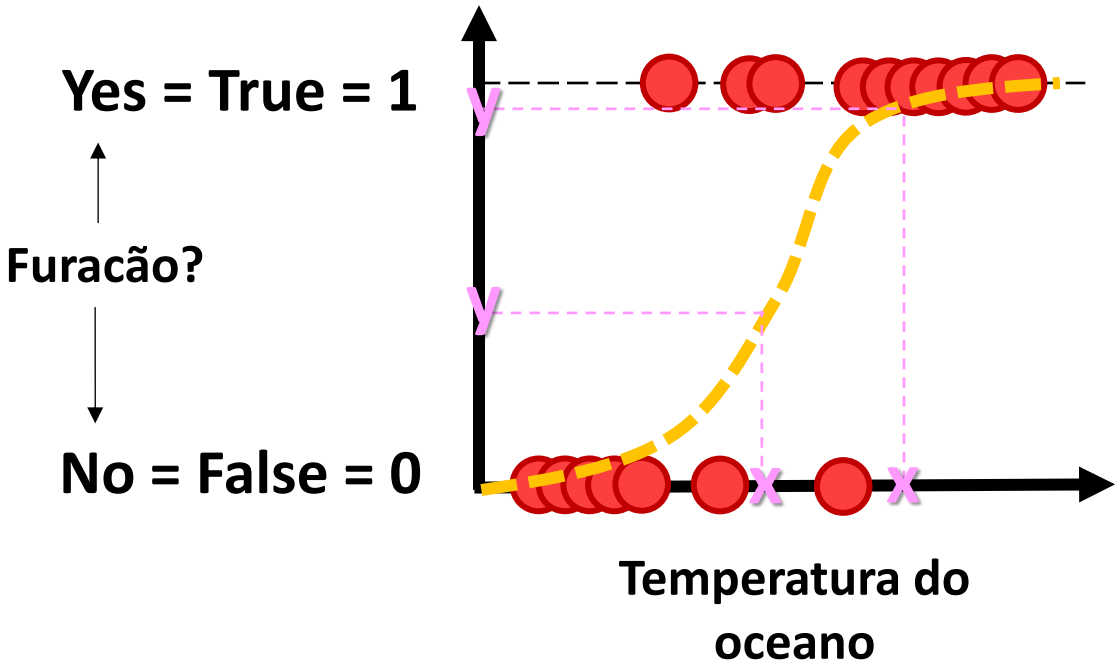


A curva vai de 0 a 1

… Isso significa que, em nosso exemplo, a curva nos diz a probabilidade de ter um furacão com base na temperatura do oceano

Ajustando um modelo de regressão logística

Podemos prever furacões com base na temperatura do oceano?

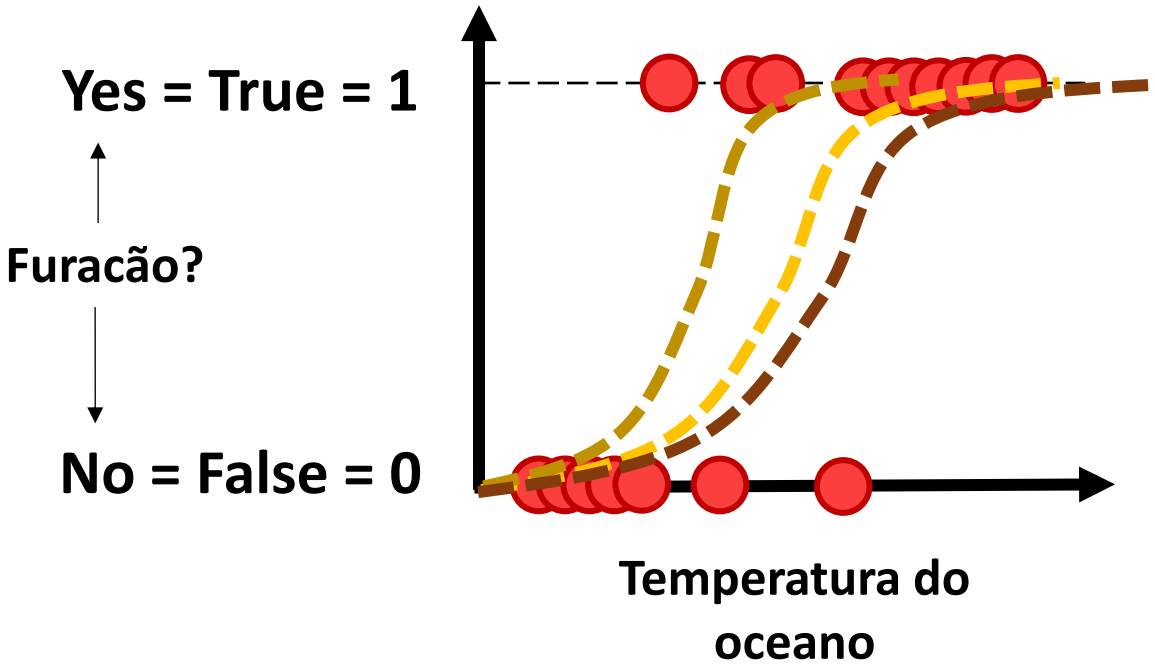


Com altas temperaturas do oceano, temos uma grande chance de ter um furacão

… Com temperaturas oceânicas mais baixas, temos uma chance menor de ter um furacão

Ajustando um modelo de regressão logística

Podemos prever furacões com base na temperatura do oceano?

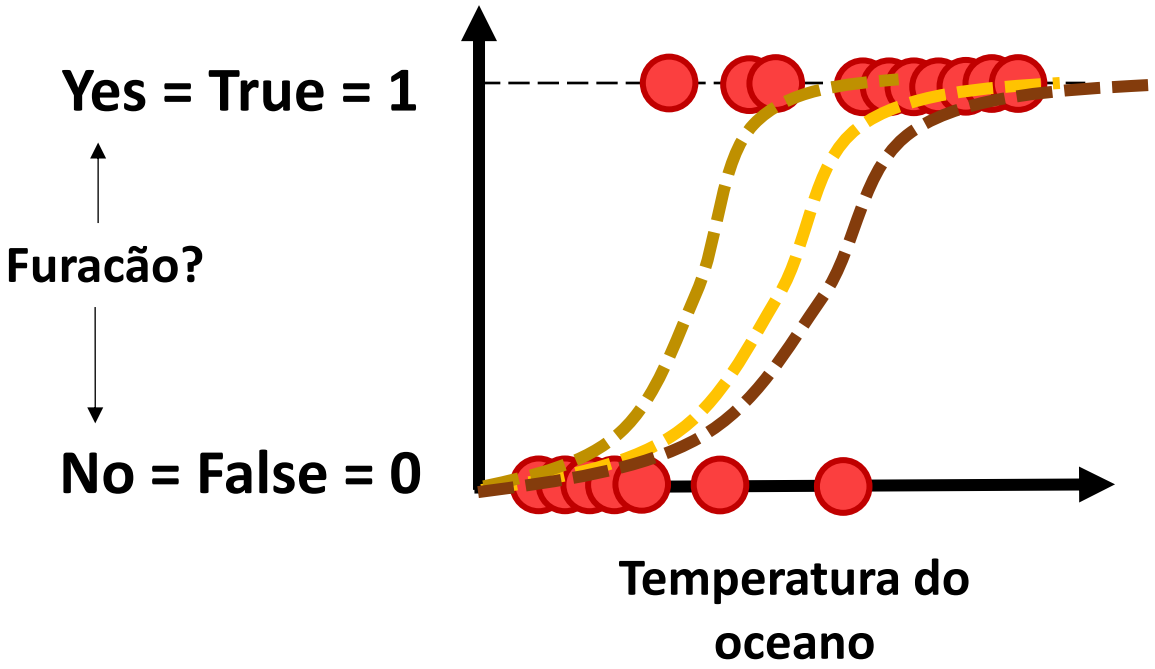


Mas como ajustar a curva/identificar qual curva é a melhor?

… Ao contrário dos modelos de regressão linear, o "melhor" ajuste para regressões logísticas não é identificado com base na menor Soma dos Resíduos Quadrados

Ajustando um modelo de regressão logística

Podemos prever furacões com base na temperatura do oceano?



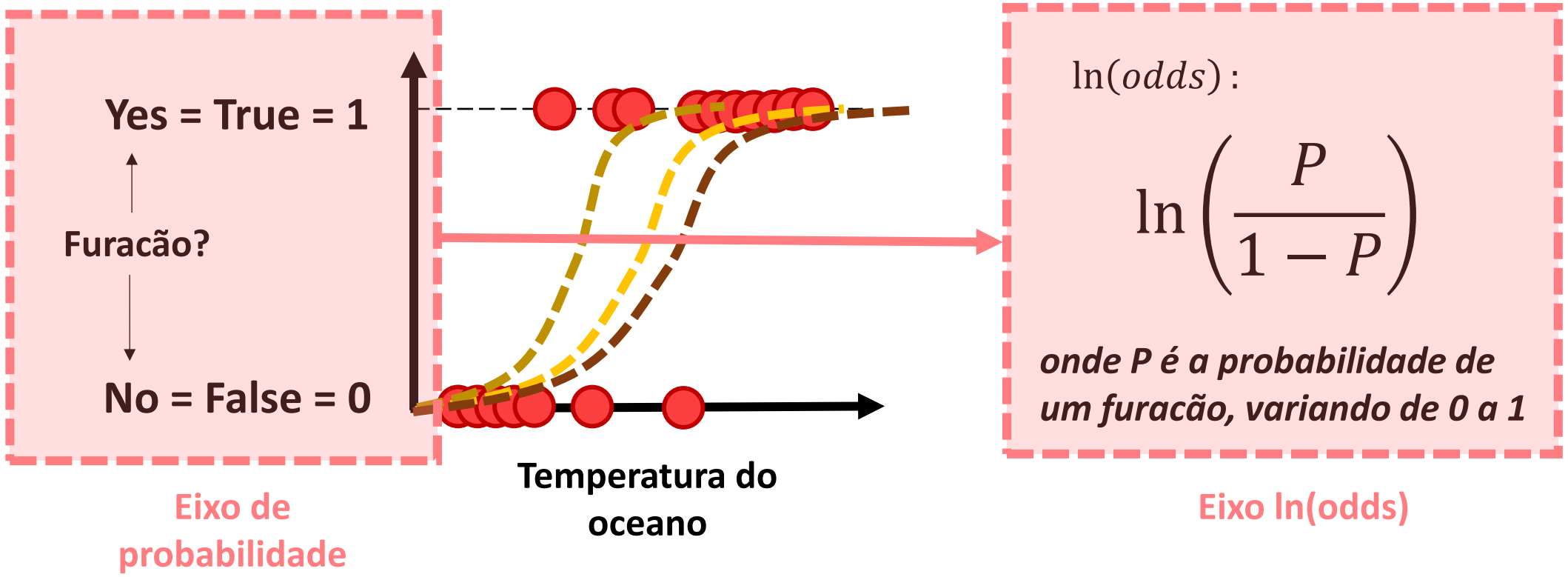
O ajuste é baseado na máxima verossimilhança (e iterativo): não existe solução em "forma fechada"

O algoritmo encontra a localização da curva em forma de S que maximiza a probabilidade de os dados observados serem gerados pelo modelo

Ajustando um modelo de regressão logística

$$\ln\left(\frac{P}{1-P}\right) = \alpha + \beta x$$

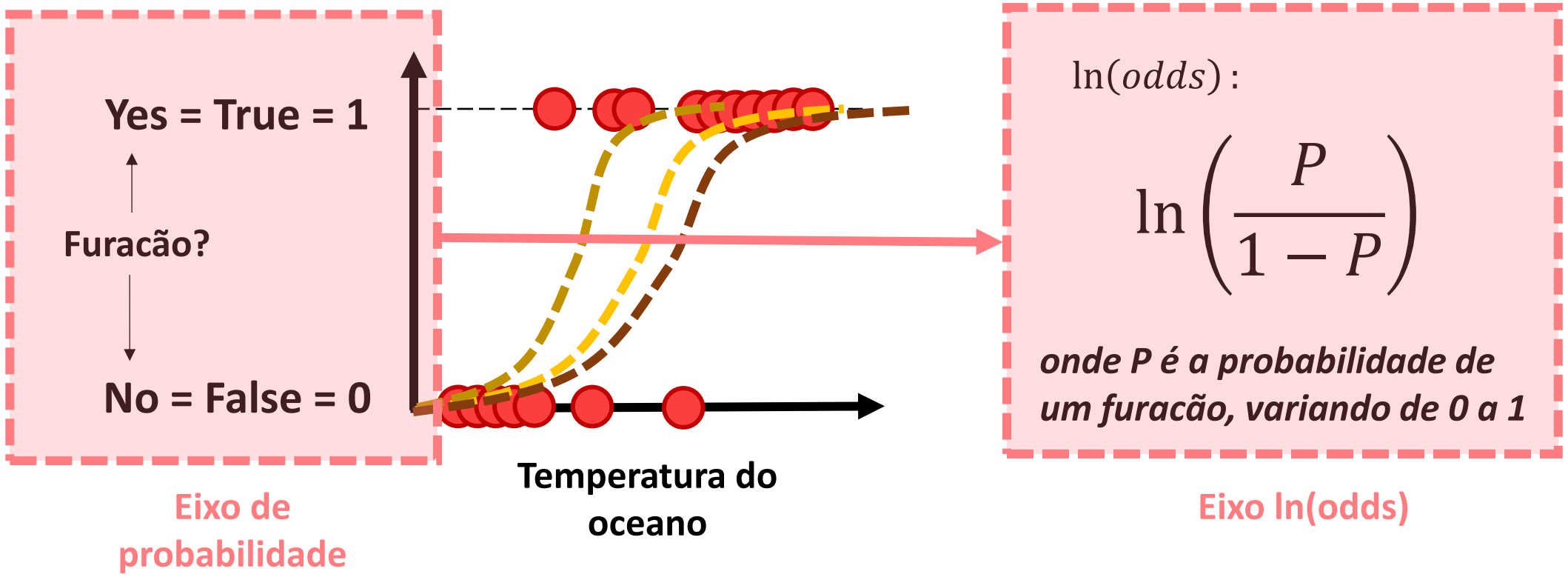
*Nas regressões logísticas, as probabilidades representam a razão entre a probabilidade de um evento acontecer e a probabilidade dele não acontecer: chamamos isso de **"odds"***



Ajustando um modelo de regressão logística

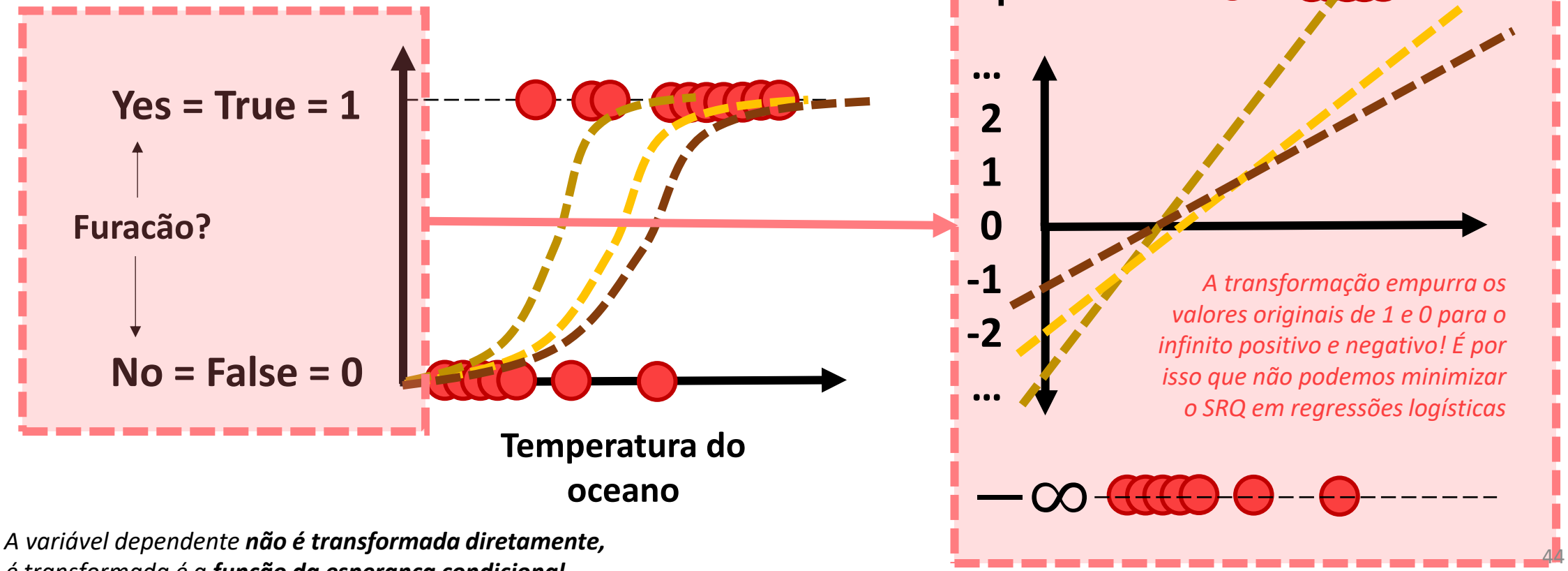
$$\ln\left(\frac{P}{1-P}\right) = \alpha + \beta x$$

*O "log-odds", também conhecido como **função logit**, é o **logaritmo natural das "odds"**. Na regressão logística, as chances logarítmicas da variável dependente são modeladas como uma combinação linear das variáveis independentes e do intercepto*



Ajustando um modelo de regressão logística

$$\ln\left(\frac{P}{1-P}\right) = \alpha + \beta x$$



*Nota: A variável dependente **não é transformada diretamente,** quem é transformada é a **função da esperança condicional***

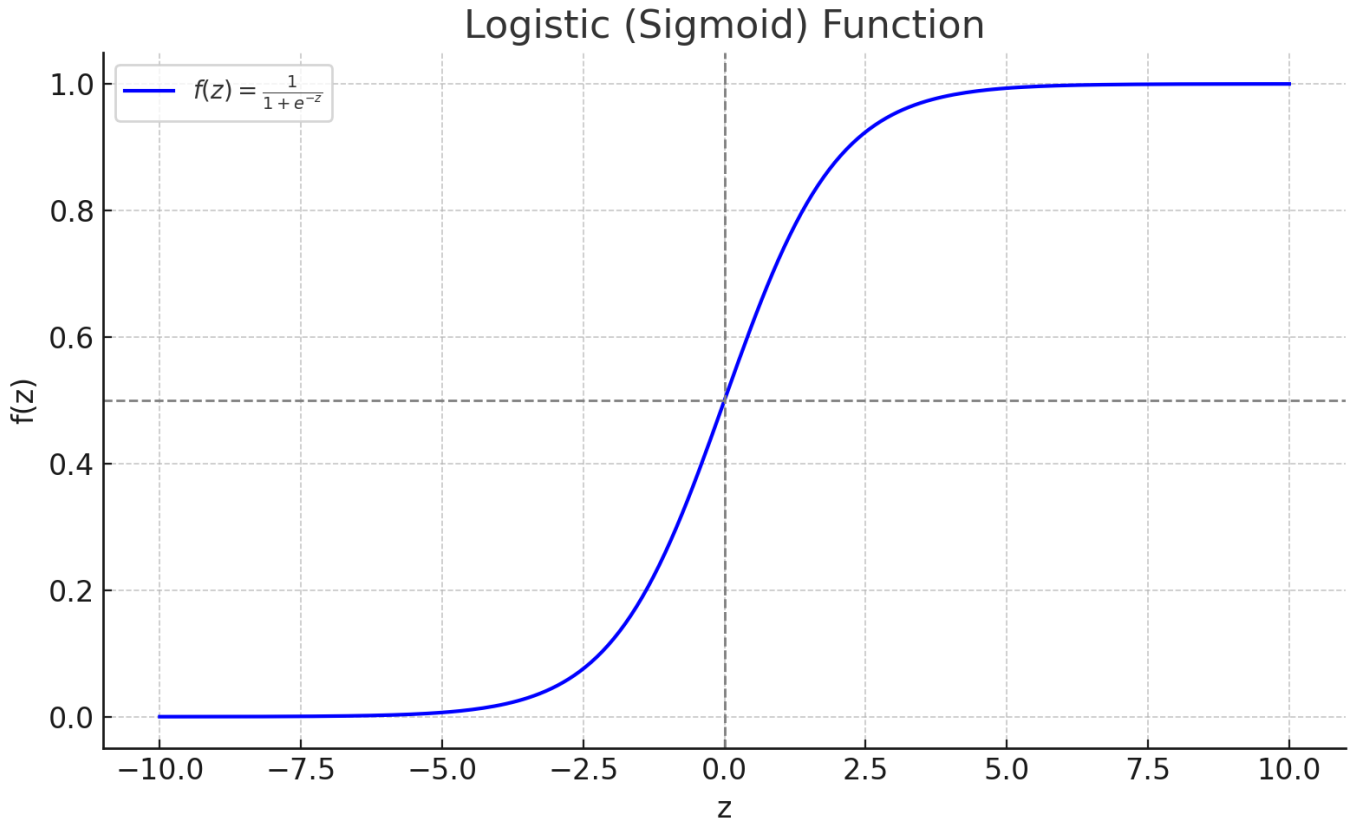
Ajustando um modelo de regressão logística

$$f(z) = \frac{1}{1 + e^{-z}}$$

$$z = \beta_0 + \beta_1 x$$

$$\hat{p} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

$$p = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$



Ajustando um modelo de regressão logística

$$Odds\ Ratio = \frac{p}{1-p}$$

$$\begin{cases} p = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} \\ 1-p = 1 - \frac{e^{(\beta_0+\beta_1 x)}}{1+e^{(\beta_0+\beta_1 x)}} = \frac{1}{1+e^{(\beta_0+\beta_1 x)}} \end{cases}$$

$$\frac{p}{1-p} = \frac{\frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}}{\frac{1}{1+e^{(\beta_0+\beta_1 x)}}} = e^{(\beta_0+\beta_1 x)}$$

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$$

Ajustando um modelo de regressão logística

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$$

Para p=1:

$$\begin{aligned}\log_e\left(\frac{p}{1-p}\right) &= \log_e(1/0)\\ &= \log_e(1) - \log_e(0)\\ &= 0 - (-\infty)\\ &= \infty\end{aligned}$$

Para p=0:

$$\begin{aligned}\log_e\left(\frac{p}{1-p}\right) &= \log_e(0/1)\\ &= \log_e(0) - \log_e(1)\\ &= (-\infty) - 0\\ &= -\infty\end{aligned}$$

Para p=0,75:

$$\begin{aligned}\log_e\left(\frac{p}{1-p}\right) &= \log_e(0.75/0.25)\\ &= \log_e(3)\\ &= 1.09\end{aligned}$$

Para p=0,3:

$$\begin{aligned}\log_e\left(\frac{p}{1-p}\right) &= \log_e(0.3/0.7)\\ &= \log_e(3) - \log_e(7)\\ &= -0.84\end{aligned}$$

Ajustando um modelo de regressão logística

https://www.geeksforgeeks.org/role-of-log-odds-in-logistic-regression/

Suposições do modelo de regressão logística

Observações independentes: *Cada observação é independente da outra, o que significa que não há correlação entre nenhuma variável de entrada*

Variáveis dependentes binárias: *Assume que a variável dependente deve ser binária ou dicotômica, o que significa que pode assumir apenas dois valores*

Relação de linearidade entre variáveis independentes e probabilidades logarítmicas: *A relação entre as variáveis independentes e as probabilidades logarítmicas da variável dependente deve ser linear*

Sem valores discrepantes: *não deve haver valores discrepantes no conjunto de dados*

Tamanho da amostra grande: *o tamanho da amostra é suficientemente grande*

Modelos de regressão logística para classificação

$$\ln\left(\frac{P}{1-P}\right) = \alpha + \beta x$$

- Se p ≥ 0,5, classificar como classe 1
- Se p < 0,5, classificar como classe 0

O limite de probabilidade usado para classificação na regressão logística é normalmente definido em 0,5 por padrão. Isso significa que, se a probabilidade prevista de uma observação pertencente à categoria positiva (classe 1) for 0,5 ou maior, a observação será classificada nessa categoria; caso contrário, é classificado na categoria negativa (classe 0)

→ *Vamos explorar outros limites quando falarmos sobre ROC/AUC*

Agenda

Parte 1. Modelos de regressão linear

Parte 2. Modelos de regressão logística

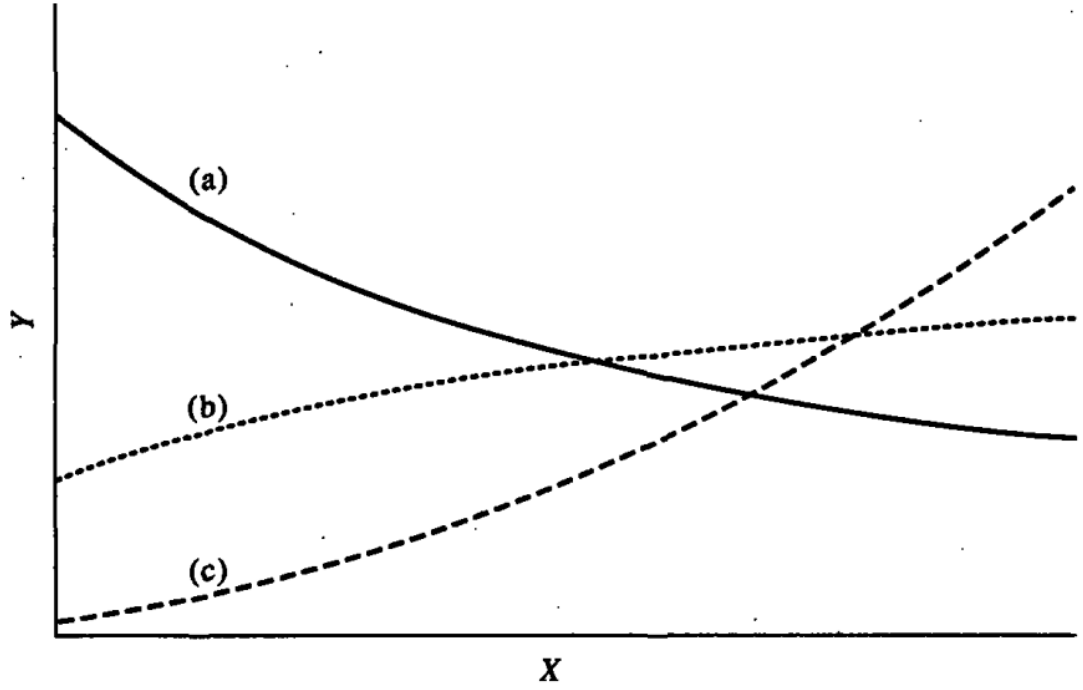
Parte 3. Modelos de regressão não-linear

Existem muitos modelos de regressão não-linear

Modelos de regressão não-linear podem ser agrupados em diferentes famílias com base em suas formas e suposições...

FIGURE 9.2.1

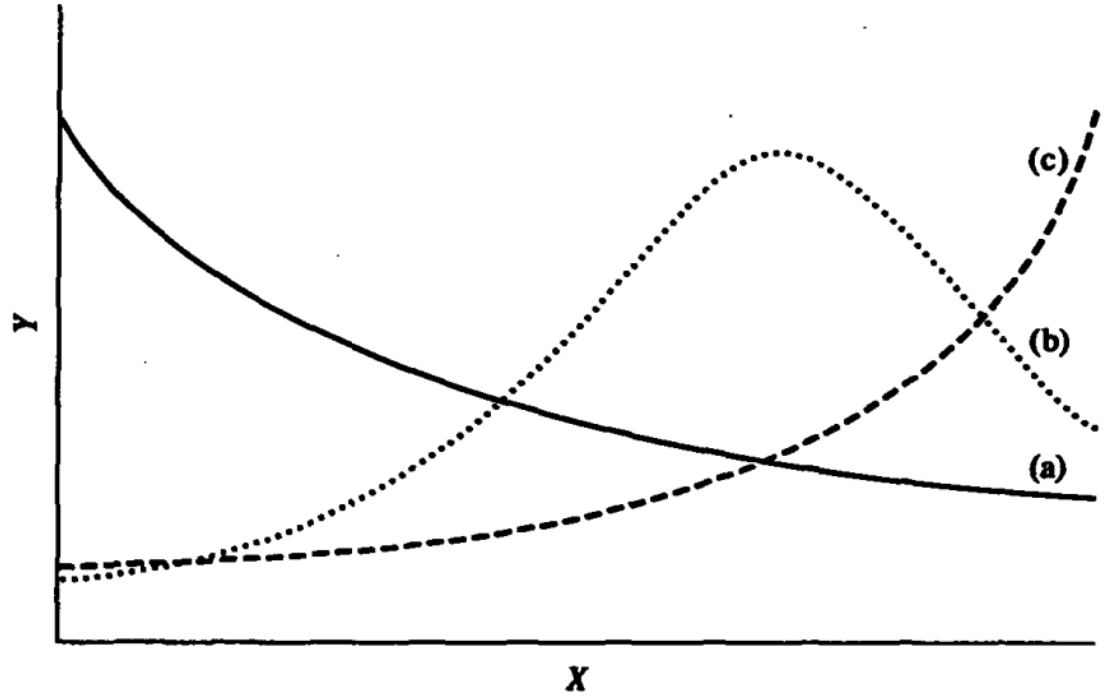
Three members of the family of curves $\mu_Y(x) = \frac{1}{(\beta_1+\beta_2 x)^{\beta_3}}$



(a) $\beta_1 = 0.6, \beta_2 = 1, \beta_3 = -1$; (b) $\beta_1 = 0.1, \beta_2 = 1, \beta_3 = 0.3$;
(c) $\beta_1 = 0.2, \beta_2 = 1, \beta_3 = 2$.

FIGURE 9.2.2

Three members of the family of curves $\mu_Y(x) = \frac{1}{\beta_1+\beta_2 x+\beta_3 x^2}$



(a) $\beta_1 = 1, \beta_2 = 3, \beta_3 = -0.2$; (b) $\beta_1 = 8.94, \beta_2 = -22.4, \beta_3 = 16$;
(c) $\beta_1 = 8, \beta_2 = -8, \beta_3 = 1$.

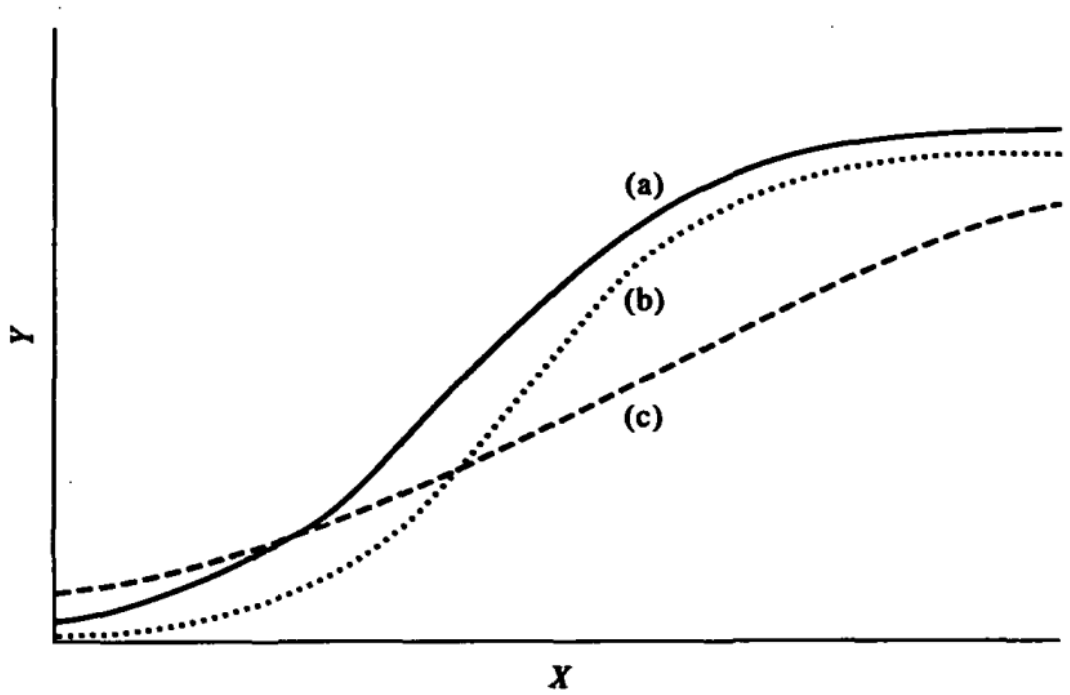
https://www.stat.colostate.edu/regression_book/chapter9.pdf

Existem muitos modelos de regressão não-linear

Modelos de regressão não-linear podem ser agrupados em diferentes famílias com base em suas formas e suposições...

FIGURE 9.2.5

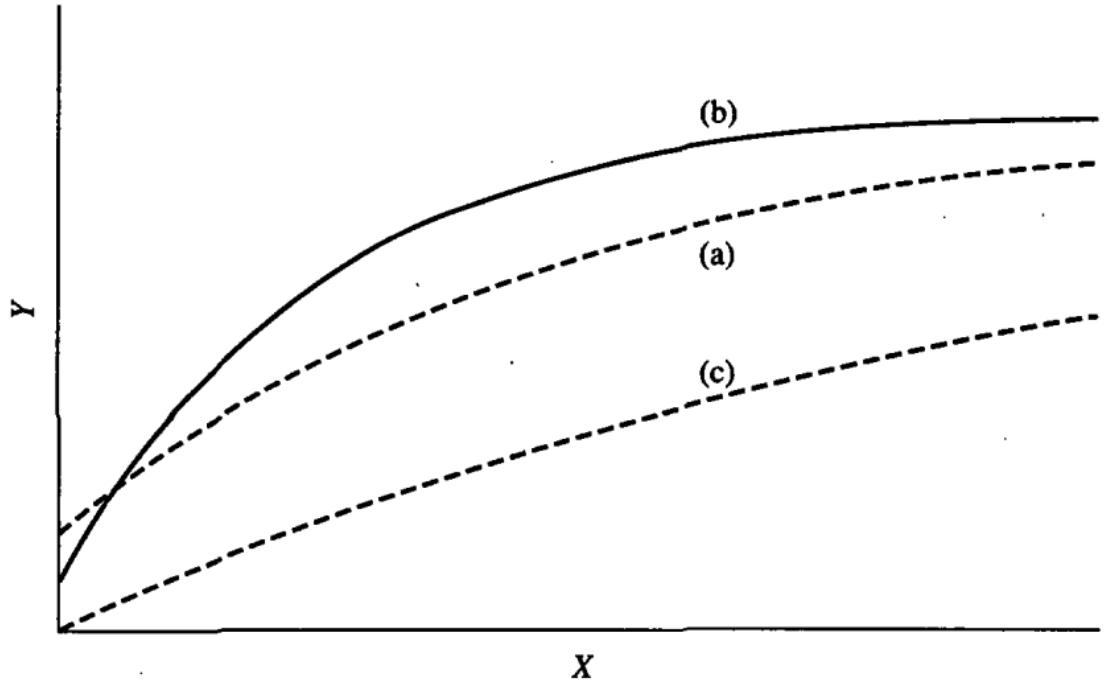
Three members of the family of curves $\mu_Y(x) = \frac{\beta_1}{1+e^{-(\beta_2+\beta_3 x)}}$



(a) $\beta_1 = 1, \beta_2 = -3.22, \beta_3 = 8$;
(b) $\beta_1 = 0.95, \beta_2 = -4.61, \beta_3 = 10$;
(c) $\beta_1 = 1, \beta_2 = -2.3, \beta_3 = 4$.

FIGURE 9.2.7

Three members of the family of curves $\mu_Y(x) = \beta_1 + \beta_2 e^{-\beta_3 x}$



(a) $\beta_1 = 1, \beta_2 = -0.9, \beta_3 = 4$; (b) $\beta_1 = 1, \beta_2 = -0.8, \beta_3 = 2$;
(c) $\beta_1 = 1, \beta_2 = -1, \beta_3 = 0, 9$.

https://www.stat.colostate.edu/regression_book/chapter9.pdf

Existem muitos modelos de regressão não-linear

https://doc.curveexpert.net/curveexpert/basicv1/modelsbuiltin.html

CurveExpert Basic v1.4 documentation » PDF Manual » Epub Manual »

Appendix: Built-In Nonlinear Regression Models

Nonlinear regression models come in all shapes and sizes - in CurveExpert, an attempt has been made to include as many models as possible, while at the same time providing relevant models to a wide range of applications. Also, keep in mind that if the model that you desire is not built-in to CurveExpert, you may define your own using the Define User Models facility. In CurveExpert, the nonlinear models has been divided into families based on their characteristic behavior. These families and their members are enumerated below:

Exponential Family

Exponential models have the exponential or logarithmic functions involved. They are generally convex or concave curves, but some models in this group are able to have an inflection point and a maximum or minimum.

- Exponential y=a*exp(b*x)
- Modified Exponential y = a*exp(b/x)
- Logarithm y = a+b*ln(x)
- Reciprocal Logarithm y = 1/(a+b*ln(x))
- Vapor Pressure Model y = exp(a+b/x+c*ln(x))

Power Family

The Power Family involves raising one or more parameters to the power of the independent variable, or raising the dependent variable to the power of a given parameter. This family is generally a set of convex or concave curves with no inflection points or maxima/minima.

- Power Fit y = a*x^b
- Modified Power y = a*b^x
- Shifted Power y = a*(x-b)^c

Resumo

- ***Em regressões lineares e logísticas, parâmetros estimados são fáceis de interpretar (não "caixas pretas")***
- ***Eles podem ser usados para identificar fatores que influenciam um processo, fazer previsões e estimar probabilidades***
- ***Esses modelos envolvem suposições estatísticas que não devem ser violadas (mas as vezes são...)***
- ***Existem muitos modelos/famílias de regressões não lineares por aí...***