

PREVISÃO DE SÉRIES TEMPORAIS

TÓPICOS

R

Análise exploratória

Pré-processamento:

- Balanceamento
- Outliers
- Missing values
- Normalização
- Seleção de atributos (Filtros, Wrappers, PCA)

Associação:

- Apriori
- FP-Growth
- Eclat

Classificação:

- Regressão logística
- Support Vector Machine (SVM)
- Árvores de Decisão
- Random Forest
- ~~Redes Neurais~~
- K nearest neighbors

Regressão:

- Regressão linear simples
- Regressão linear múltipla
- Regressão não linear simples
- Regressão não linear múltipla

Agrupamento:

- Particionamento (K-means, K-medoids)
- Hierárquico (DIANA, AGNES)
- Densidade (DBSCAN)

Séries Temporais:

- Naive
- Média Móvel
- Amortecimento exponencial
- Auto-regressivo integrados de média móvel
- Auto regressivo não linear

Recapitulação

REGRESSÃO

REGRESSÃO

Regressão
Linear Simples

$$y = b_0 + b_1x_1$$

Regressão
Linear Múltipla

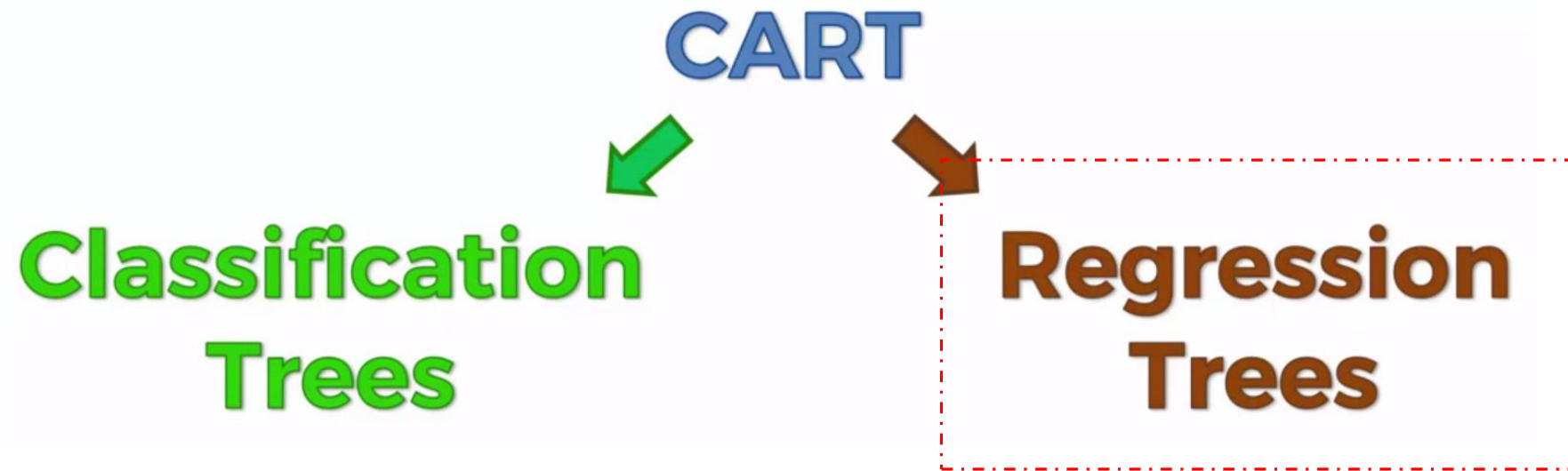
Variável dependente

Variáveis independentes

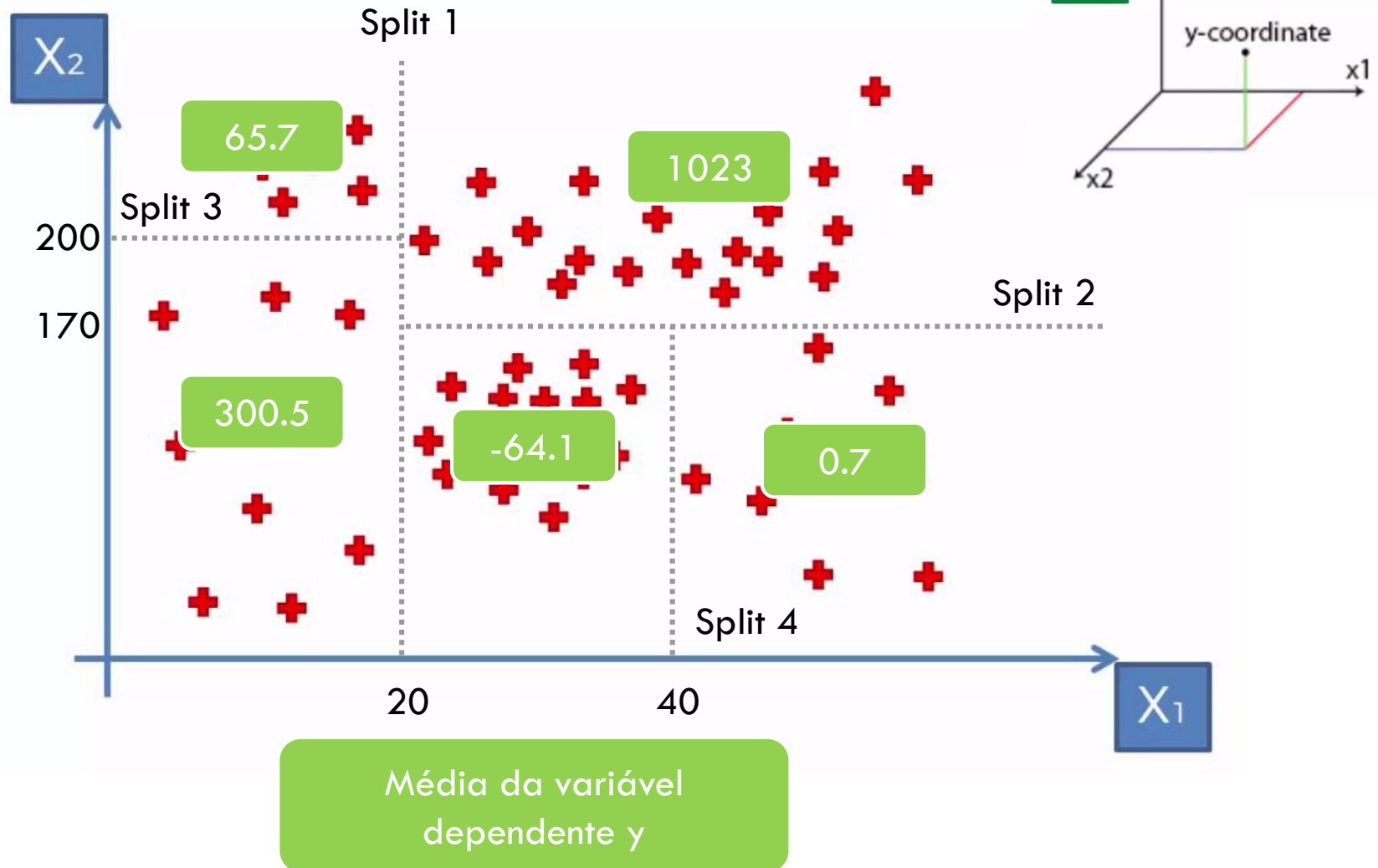
$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n$$

Árvore de Regressão

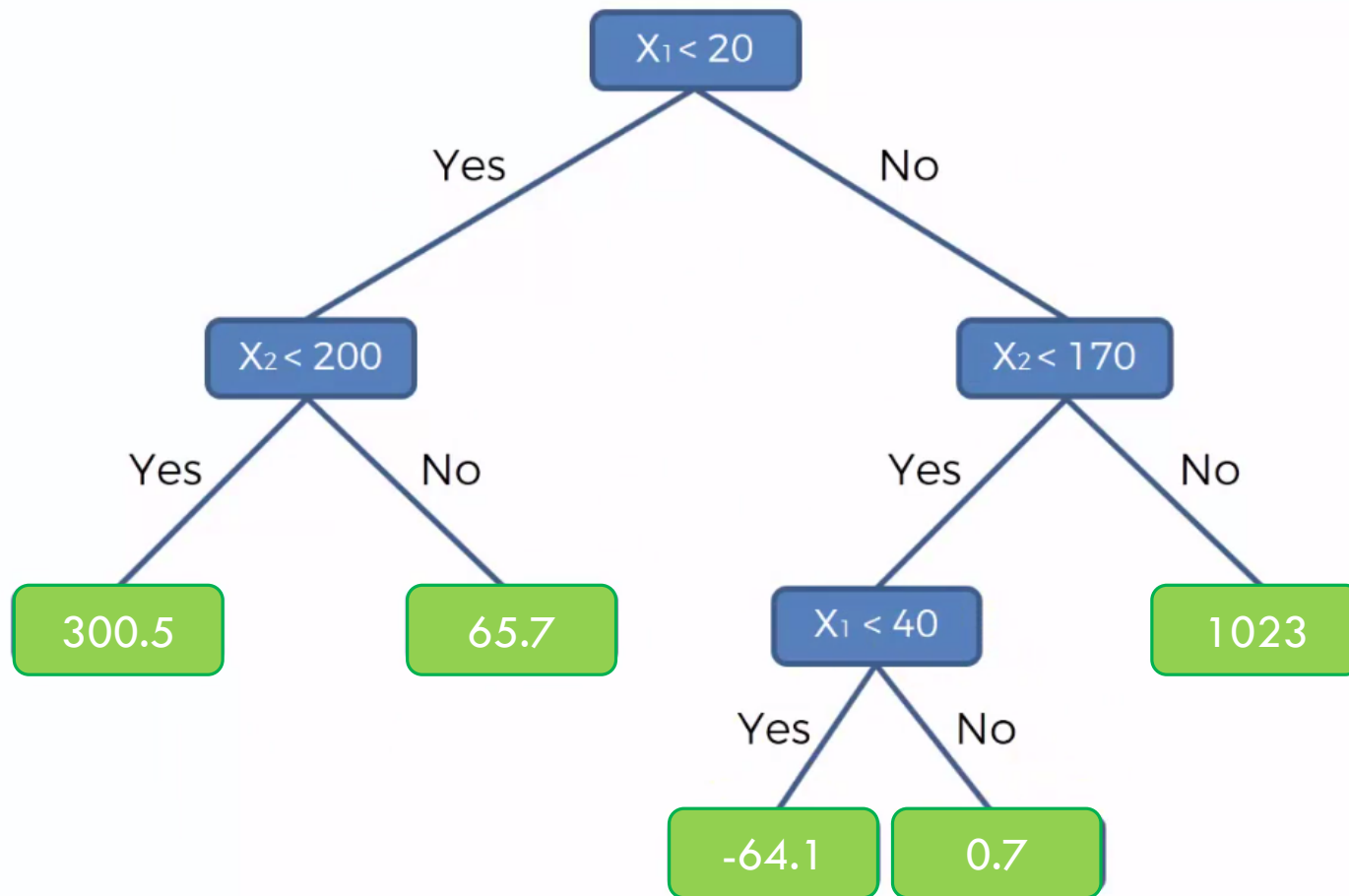
ÁRVORES DE REGRESSÃO



ÁRVORES DE REGRESSÃO

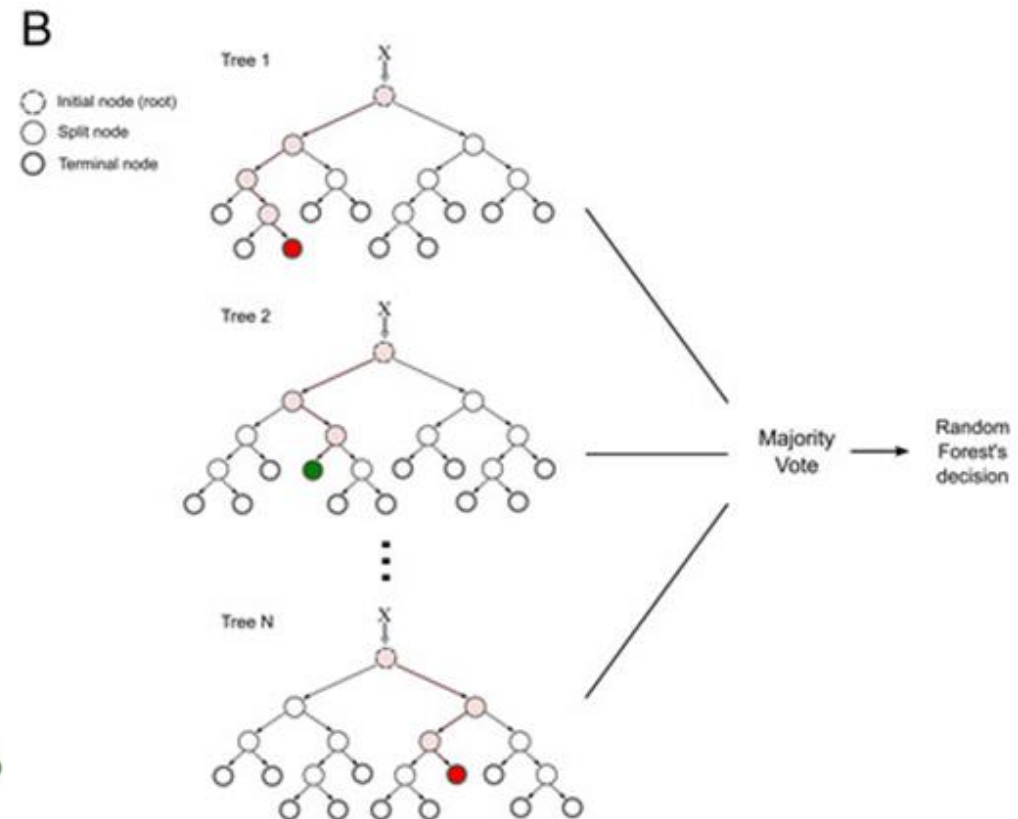
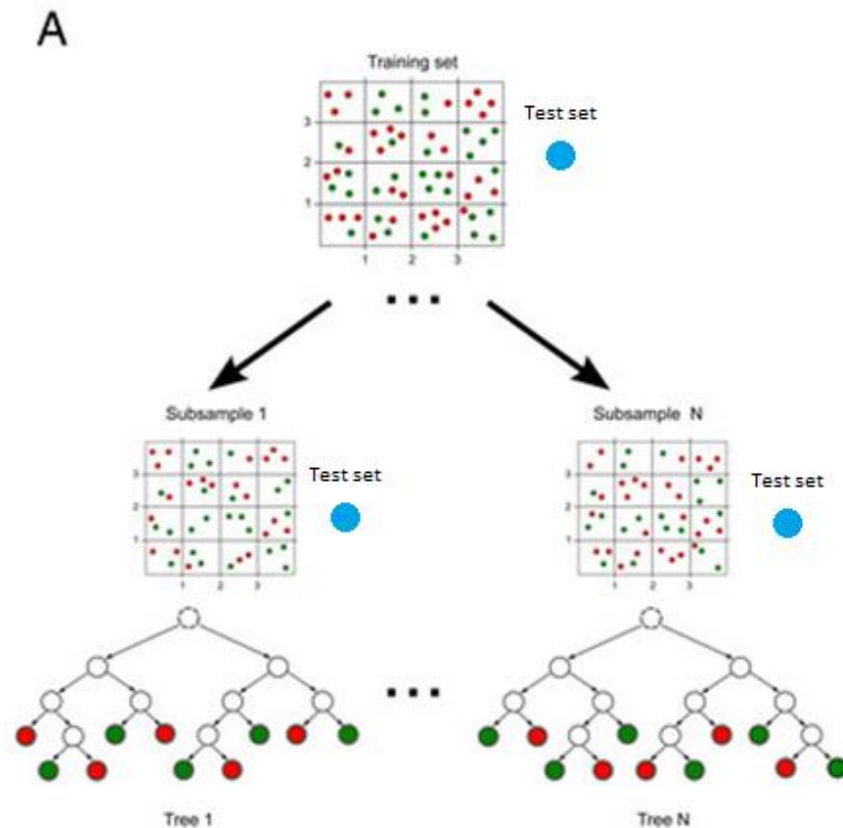


ÁRVORES DE REGRESSÃO



Random Forest

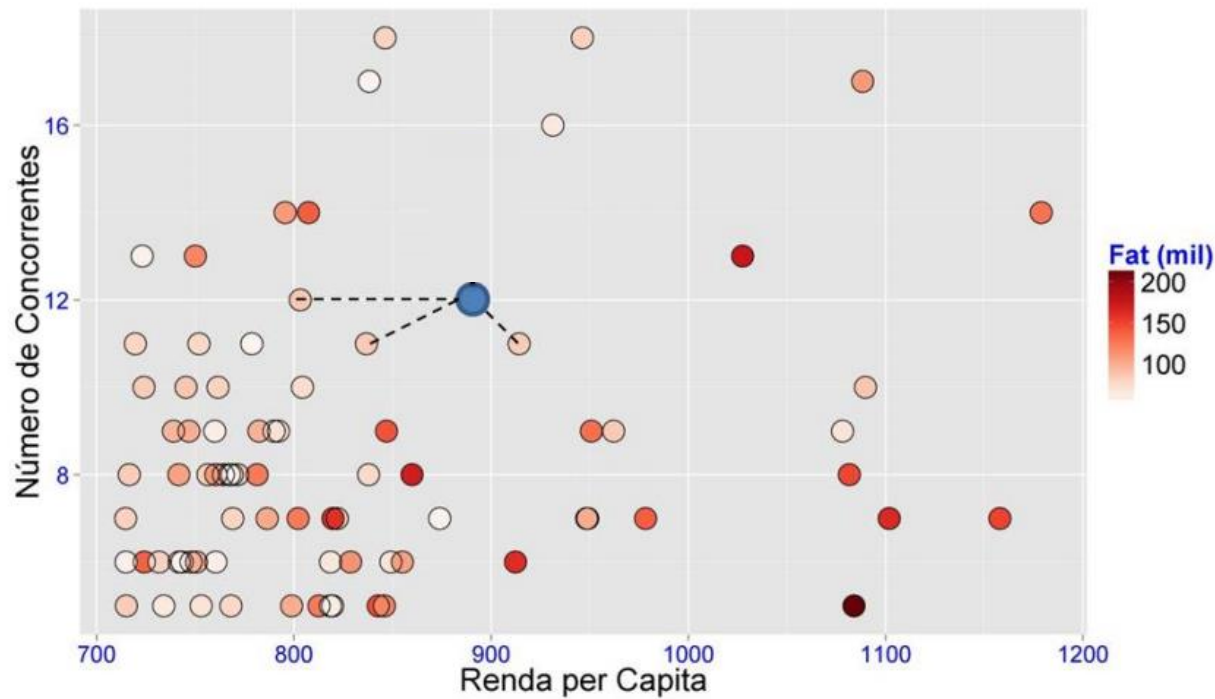
RANDOM FOREST



K Nearest Neighbors

KNN PARA REGRESSÃO

- **3NN para Regressão:** O Faturamento Esperado é a Média Aritmética dos **3** vizinhos mais próximos!



R² VERSUS R² AJUSTADO

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

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

ESTUDOS DE CASO



Previsão de Séries Temporais

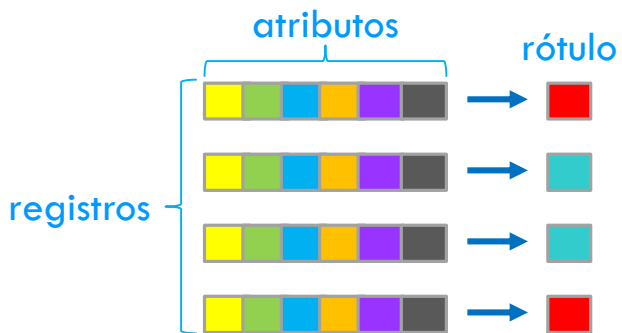
Machine Learning

Supervisionado

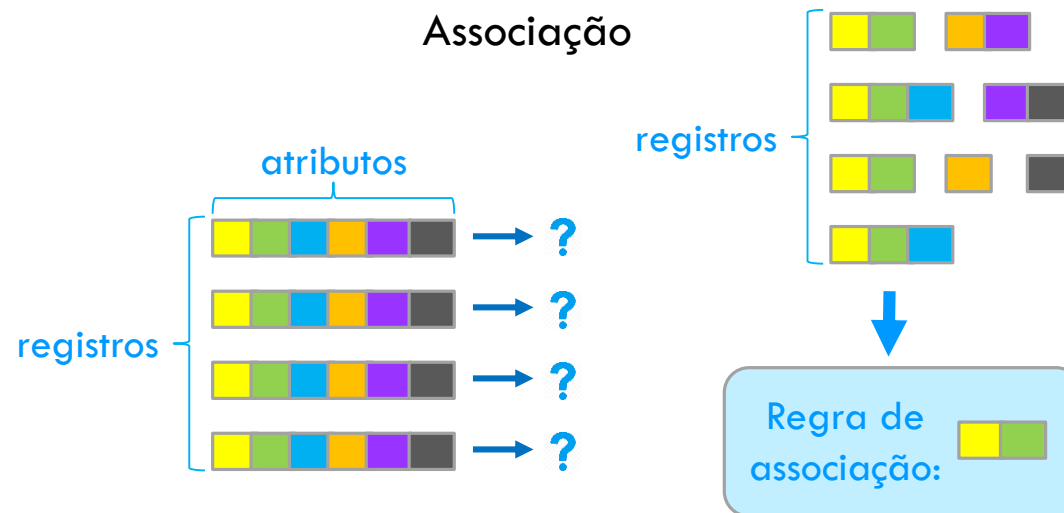
Não Supervisionado

Reforço

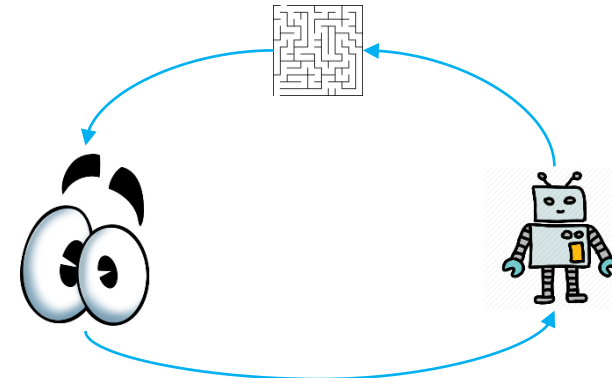
Classificação
Regressão
Previsão de séries temporais



Agrupamento
Associação



Aprendizado através da interação de agentes com um ambiente

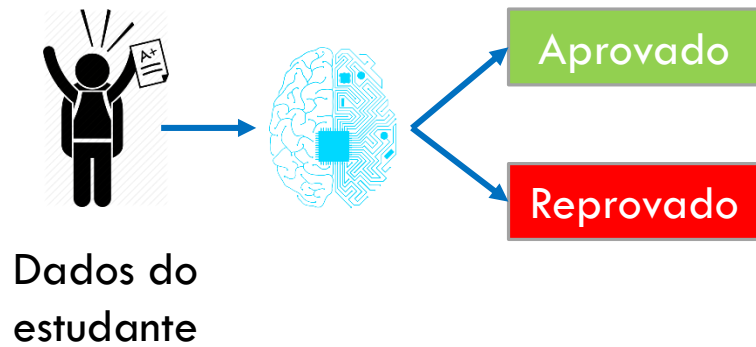


SUPERVISIONADO

- Aproximador: função mapeia entradas e saída.

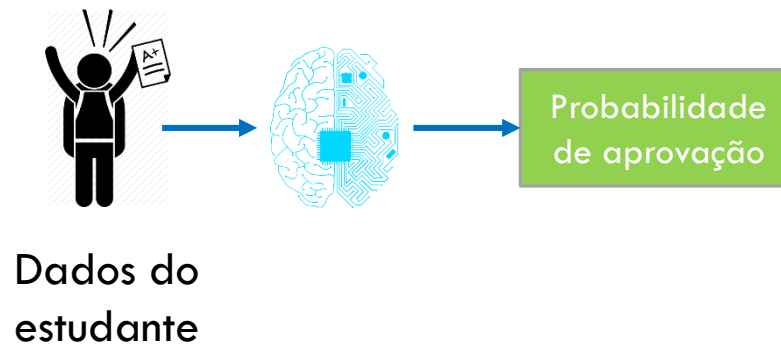
Classificação

Rótulo é categórico.



Regressão

Rótulo é contínuo.



Previsão de Séries

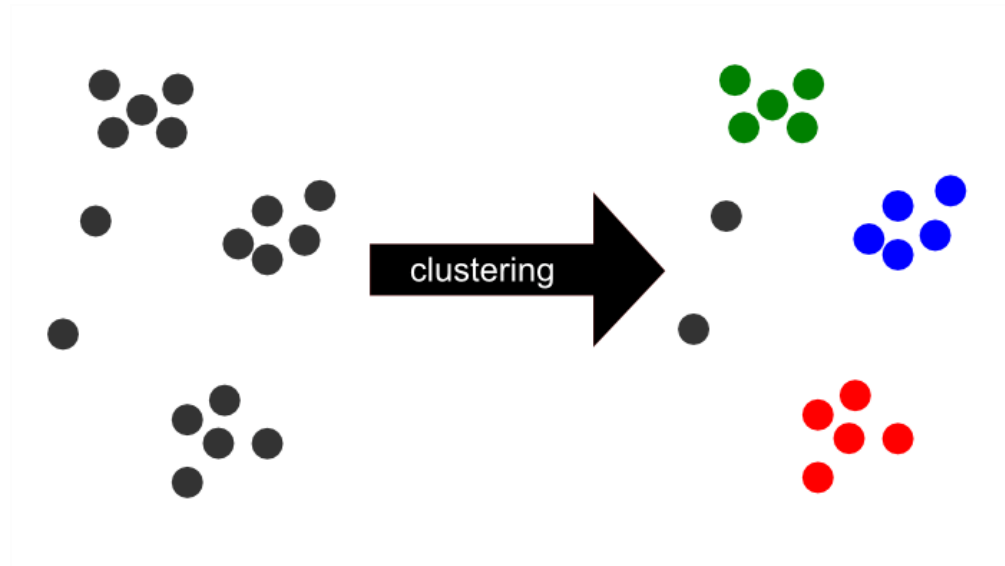
Rótulo é contínuo e dependente do tempo.



NÃO SUPERVISIONADO

Agrupamento

Descoberta de semelhanças e grupos entre registros.



Associação

Descoberta de relações entre variáveis.



SÉRIES TEMPORAIS

Uma série temporal é um conjunto de observações ordenadas no tempo, não necessariamente igualmente espaçadas, que apresentam dependência serial, isto é, dependência entre instantes de tempo.



SÉRIES TEMPORAIS

Uma série temporal é um conjunto de observações ordenadas no tempo, não necessariamente igualmente espaçadas, que apresentam dependência serial, isto é, dependência entre instantes de tempo.



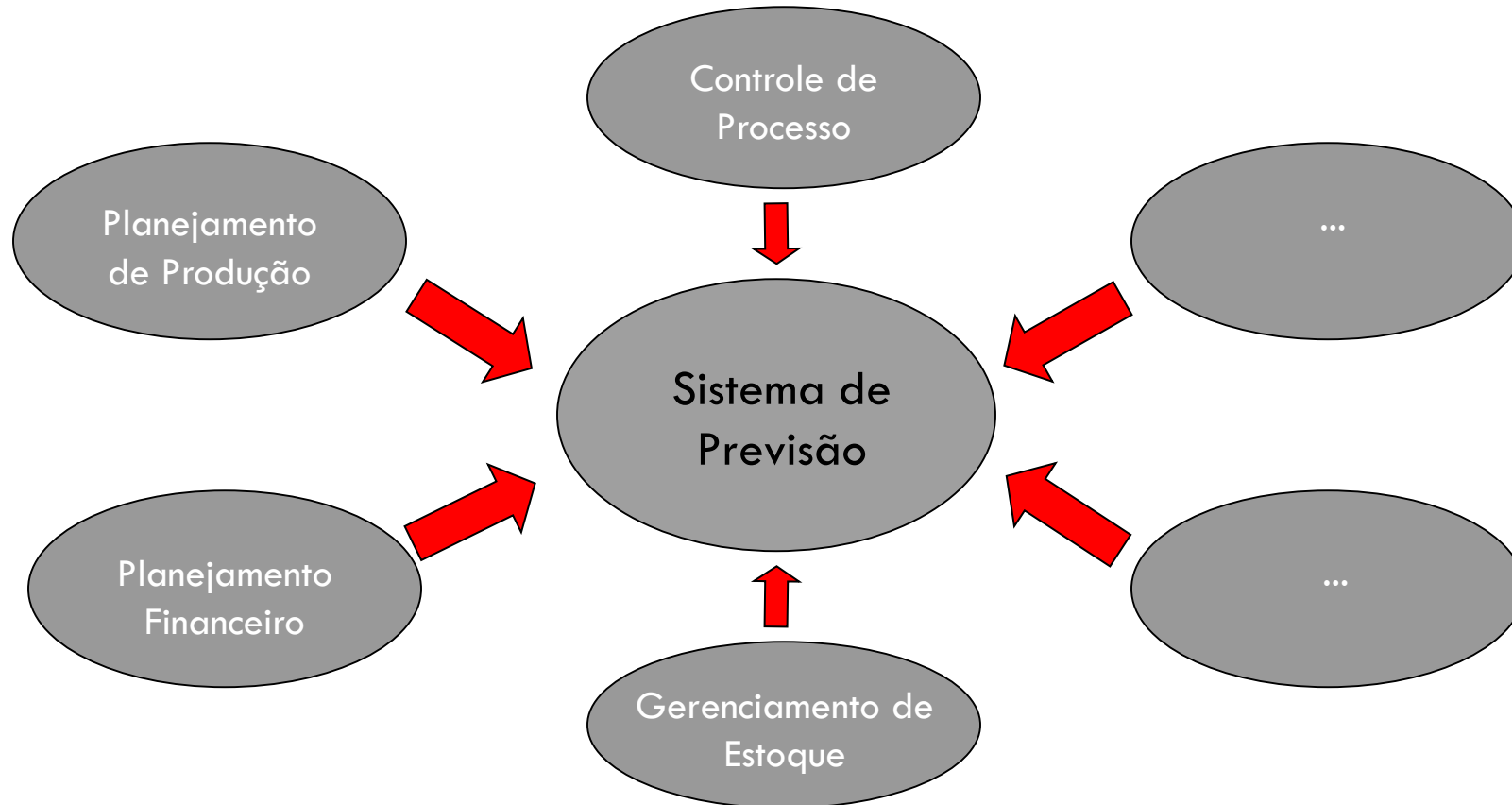
SÉRIES TEMPORAIS

Podemos enumerar os seguintes exemplos de séries temporais:

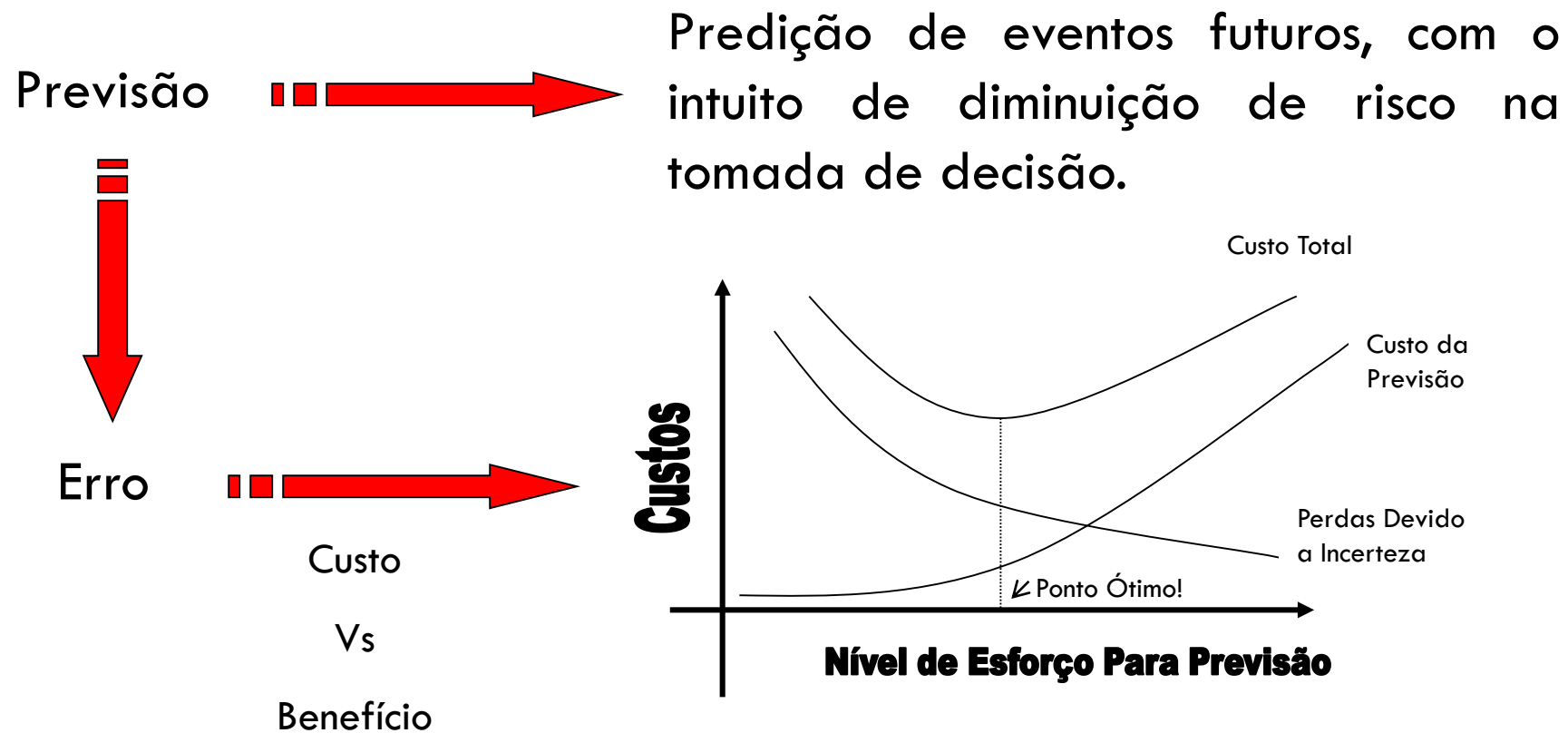
- Temperaturas máximas e mínimas diárias em uma cidade;
- Vendas mensais de uma empresa;
- Valores mensais do IPCA;
- Valores de fechamento diários do IBOVESPA.

INTRODUÇÃO

“Previsão é um elemento chave na tomada de decisão”



PREVISÃO



DECISÃO

Baseando-se em sistemas de Previsão:



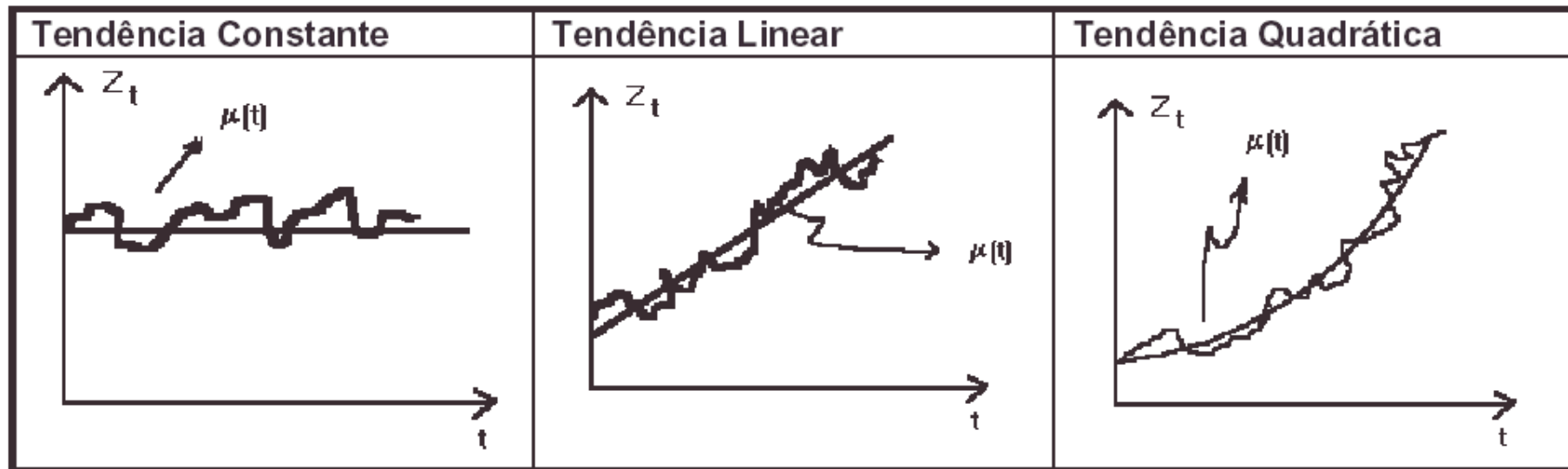
PREVISÃO DE SÉRIES TEMPORAIS

- Tipos possíveis de processamento:
- **Predição:** futuros valores de $x(t)$
- **Classificação:** uma série de uma ou algumas classes:
 - Preço vai subir
 - Preço vai cair
 - Sem mudanças
- **Transformação:** de uma série temporal em outra série temporal:
 - Preço do óleo → Consumo de gasolina nos postos

Cotação EURUSD

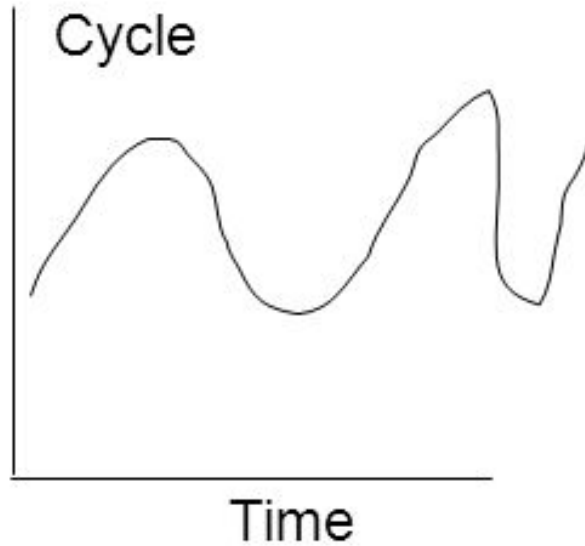
A TENDÊNCIA

Indica o seu comportamento “de longo prazo”, isto é, se ela cresce, decresce ou permanece estável, e qual a velocidade destas mudanças.



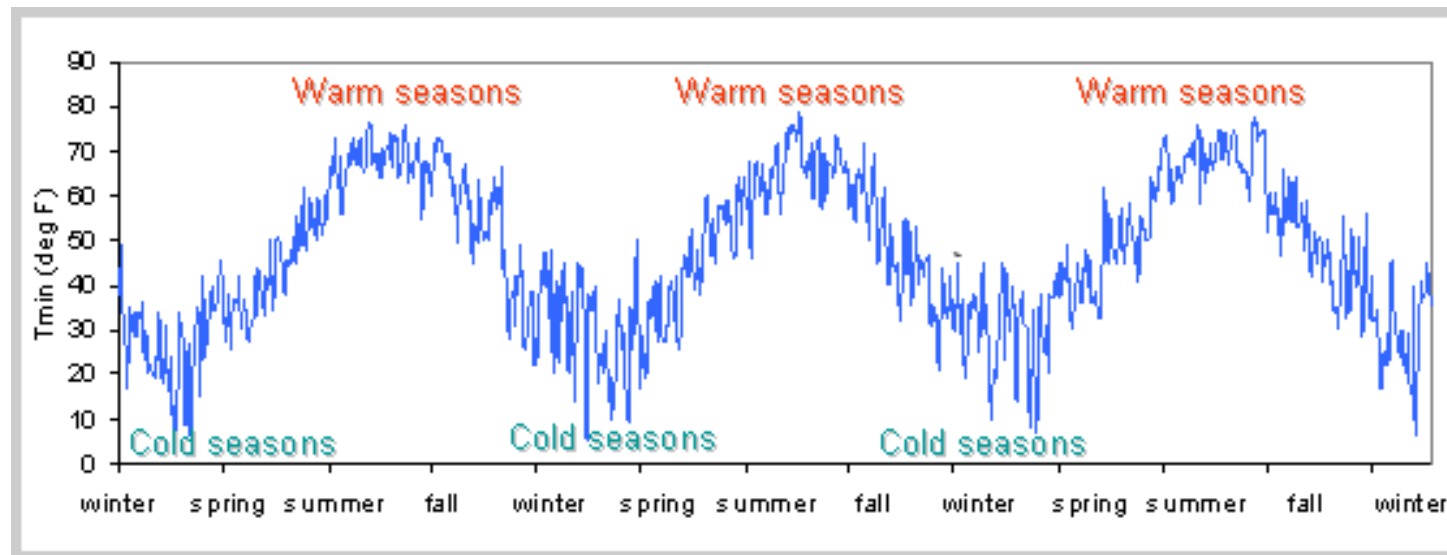
OS CICLOS

São caracterizados pelas **oscilações de subida** e de **queda** nas séries, de forma **não periódica**, ao longo da componente de tendência.



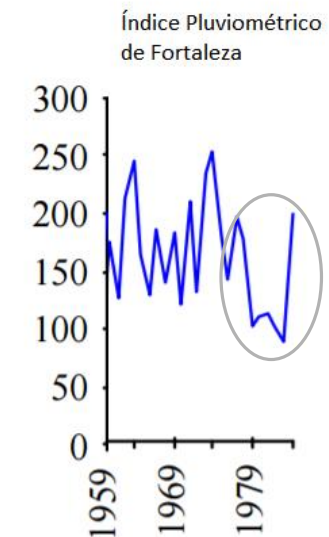
A SAZONALIDADE

Corresponde às **oscilações de subida e de queda** que sempre ocorrem em um determinado período do ano, do mês, da semana ou do dia.



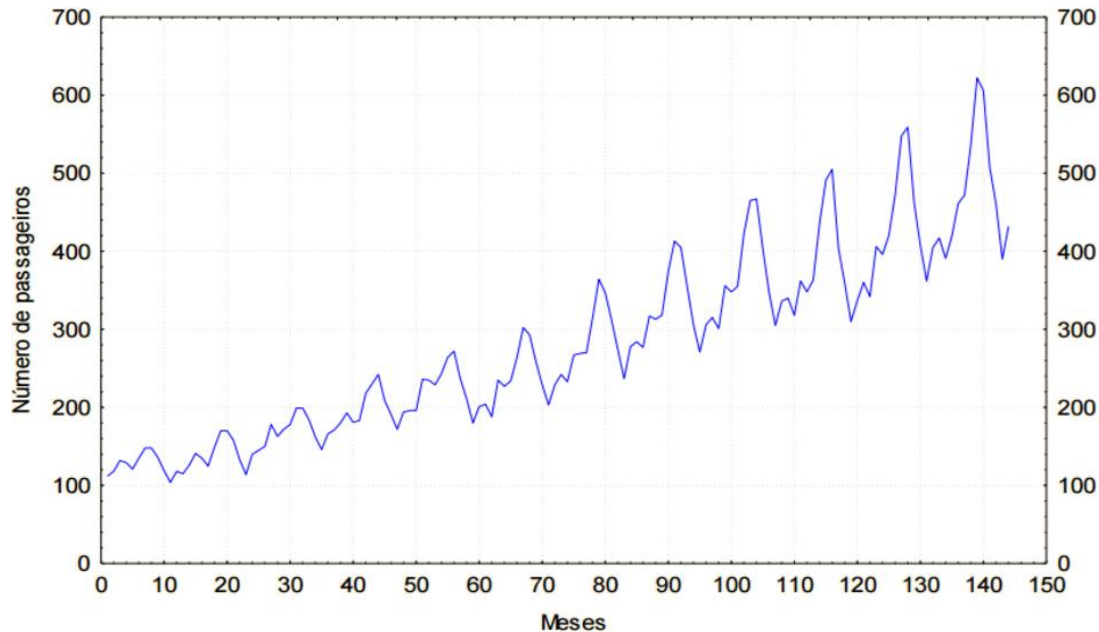
SAZONALIDADE X CICLOS

A diferença essencial entre as componentes **sazonal** e **cíclica** é que a **primeira** possui **movimentos facilmente previsíveis**, ocorrendo em **intervalos regulares de tempo (periódico)**, enquanto que movimentos cíclicos tendem a ser irregulares.



Grande seca seguida de um ano com grande quantidade de chuva.

SÉRIES TEMPORAIS

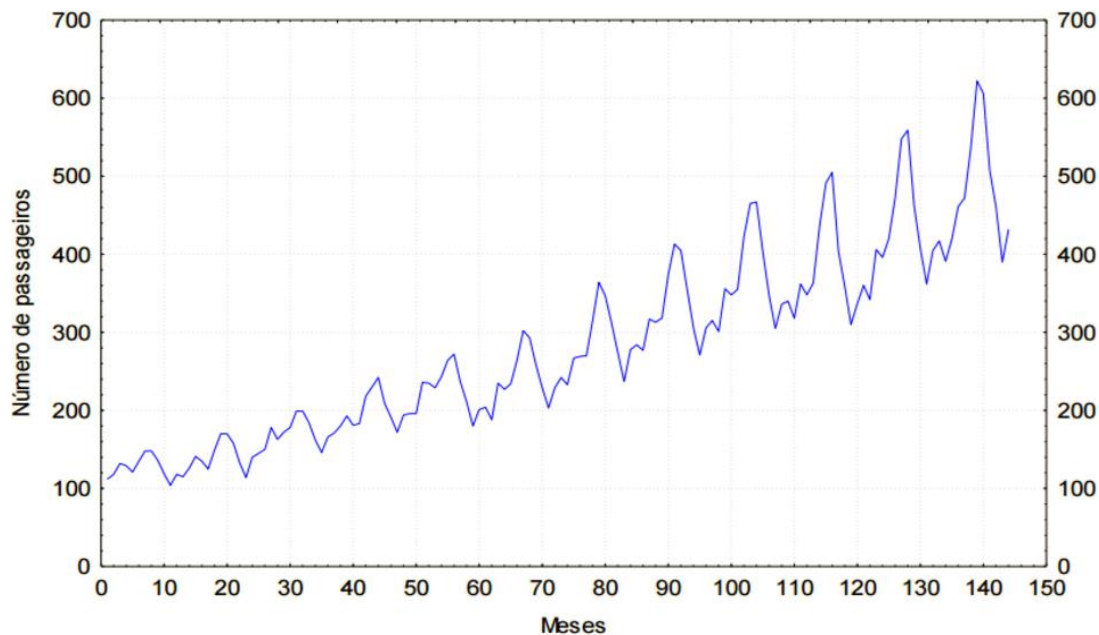


Que **padrões não aleatórios** podemos identificar na figura?

Observe que há uma **tendência crescente** no número de passageiros transportados.

Há uma **sucessão regular de "picos e vales"** no número de passageiros transportados, isso deve ser causado pelas oscilações devido a feriados, períodos de férias escolares, etc., que estão geralmente relacionados às estações do ano, e que se repetem todo ano (com maior ou menor intensidade).

SÉRIES TEMPORAIS



Que **padrões não aleatórios** podemos identificar na figura?

Em outras palavras, identificamos dois padrões que podem tornar a ocorrer no futuro:

1. crescimento no número de passageiros transportados;
2. flutuações sazonais.

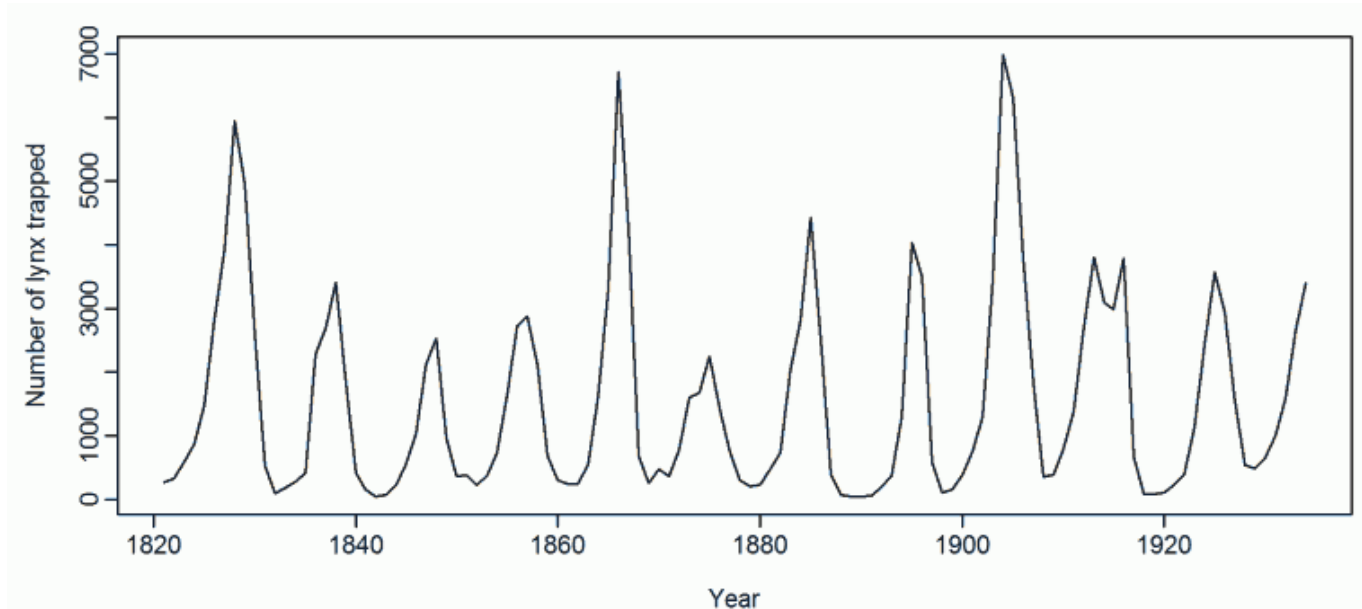
Tais padrões poderiam ser incorporados a um modelo estatístico, possibilitando fazer previsões que auxiliarão na tomada de decisões.

SÉRIES TEMPORAIS

Em geral, ao estudarmos uma série temporal estamos interessados em:

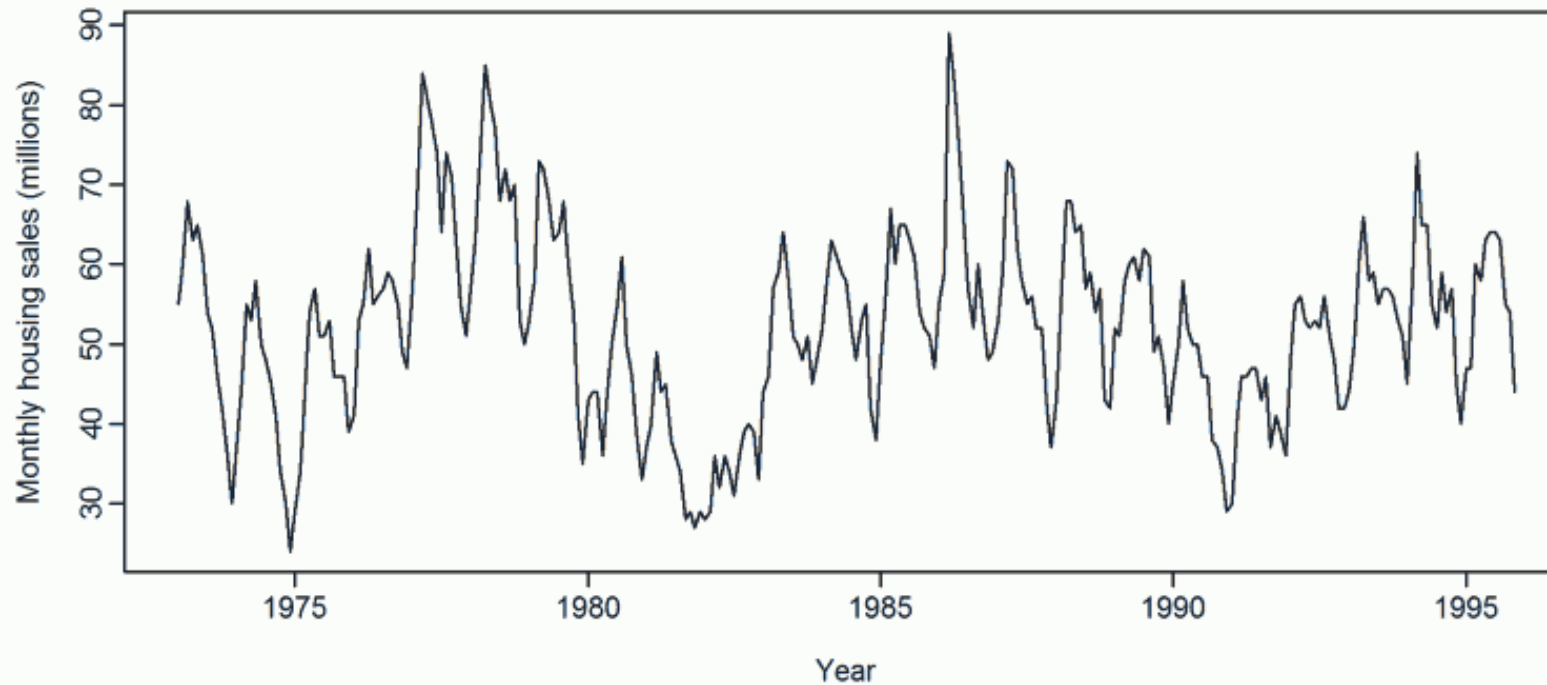
- **Análise e modelagem da série temporal** - descrever a série, verificar suas características mais relevantes e suas possíveis relações com outras séries;
- **Previsão da série temporal** - a partir de **valores históricos da série** (e possivelmente de outras séries também) **procura-se estimar previsões de curto prazo (forecast)**. O número de instantes à frente para o qual é feita a previsão é chamado de horizonte de previsão.

SÉRIES TEMPORAIS



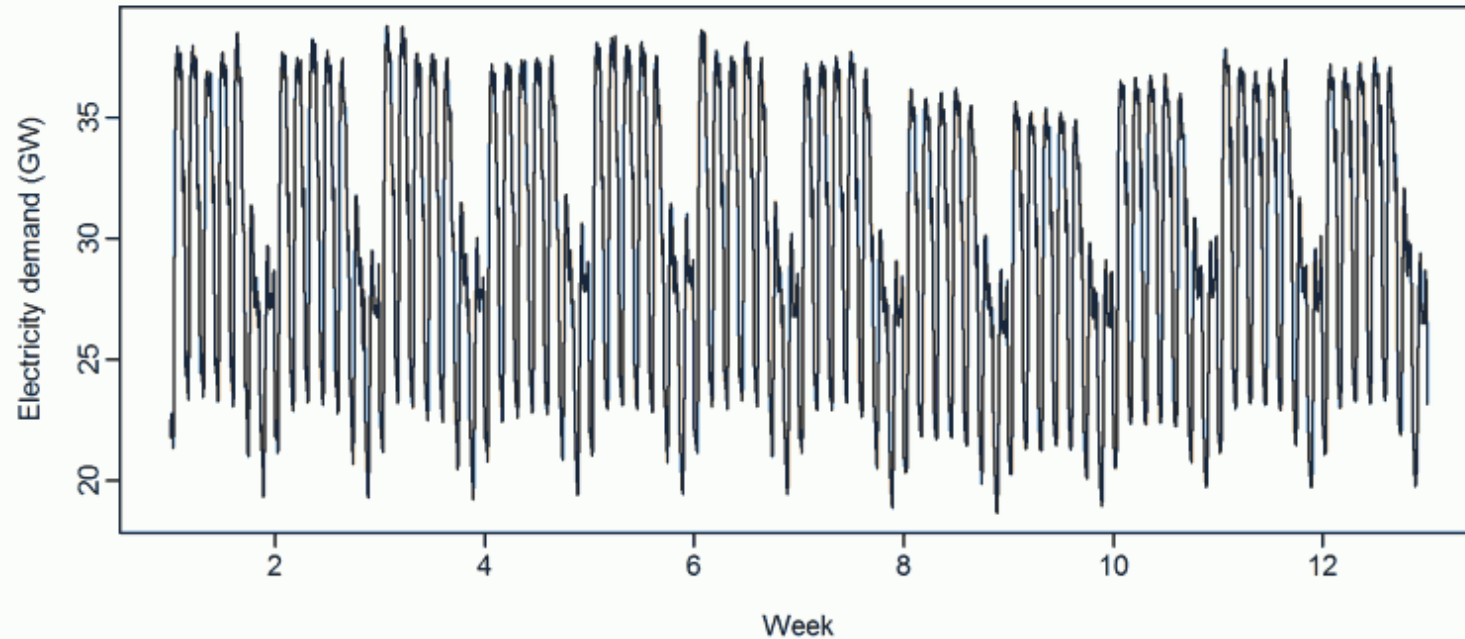
Base famosa Canadian lynx — o número de linces presos a cada ano no distrito do rio McKenzie a noroeste do Canada (1821-1934). O gráfico mostra claramente **ciclos** de população **aperiódicos** de aproximadamente 10 anos. Os ciclos não são de comprimento fixo — alguns duram 8 ou 9 anos e outros duram mais de 10 anos.

SÉRIES TEMPORAIS



Vendas mensais de casas nos Estados Unidos (1973-1995). Existe uma **sazonalidade** forte em cada ano, assim como um **comportamento cíclico** com período de duração entre 6 e 10 anos.

SÉRIES TEMPORAIS



Demanda de eletricidade na Inglaterra a cada meia hora (5 de Junho de 2000 - 27 de Agosto de 2000). Aqui existem **dois tipos de sazonalidade** — um padrão diário e um semanal.

Se houvessem dados coletados ao longo de anos, veríamos que existe também um padrão anual.

PROCEDIMENTOS ESTATÍSTICOS DE PREVISÃO

Na prática variam muito, podendo ser:

- Simples e intuitivos, com pouca análise dos dados
- Complexos e racionais, envolvendo uma considerável trabalho de interpretação de séries temporais.

A previsão não constitui um fim em si, mas deve ser vista como parte integrante de um complexo processo de tomada de decisão, visando a objetivos específicos.

MODELOS

Univariado: inclui os modelos que se baseiam em uma única série histórica. Como exemplo, podem ser citados:

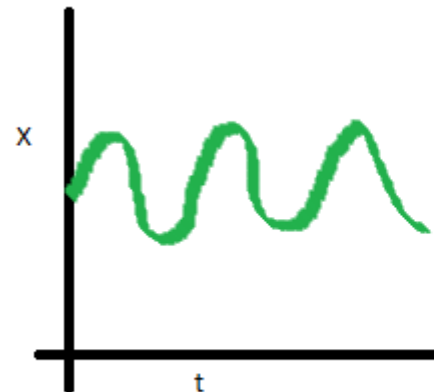
Multivariado: Modelam simultaneamente duas ou mais séries temporais sem qualquer exigência em relação à direção da causalidade entre elas.

Função de Transferência: a série de interesse é explicada não só pelo seu passado histórico, como também por outras séries temporais não correlacionadas entre si.

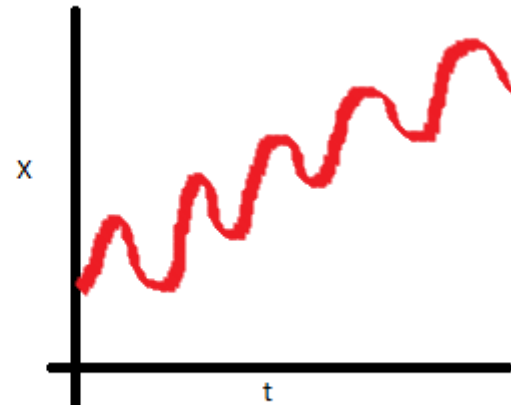
ESTACIONARIEDADE DE UMA SÉRIE

A estacionariedade em uma série temporal significa que os dados oscilam sobre uma média constante, independente do tempo, com a variância das flutuações permanecendo essencialmente a mesma.

- Média do processo é constante;



Stationary series

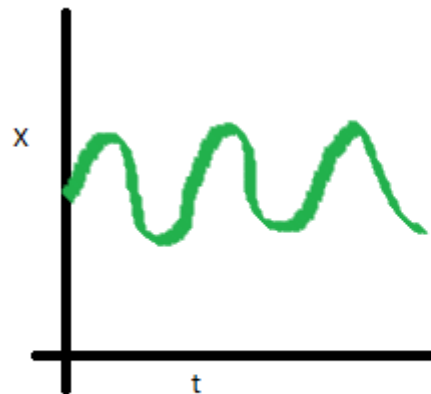


Non-Stationary series

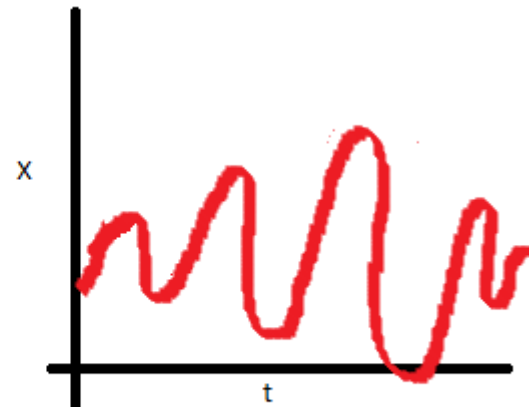
ESTACIONARIEDADE DE UMA SÉRIE

A estacionariedade em uma série temporal significa que os dados oscilam sobre uma média constante, independente do tempo, com a variância das flutuações permanecendo essencialmente a mesma.

- Média do processo é constante;
- Variância do processo é constante;



Stationary series



Non-Stationary series

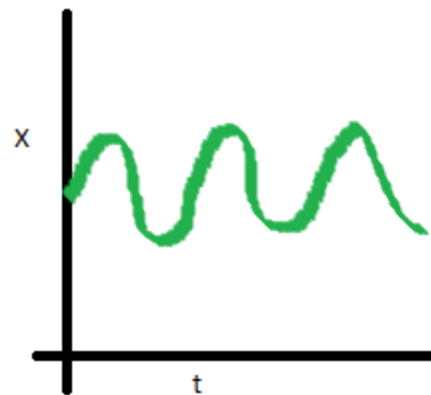
ESTACIONARIEDADE DE UMA SÉRIE

A estacionariedade em uma série temporal significa que os dados oscilam sobre uma média constante, independente do tempo, com a variância das flutuações permanecendo essencialmente a mesma.

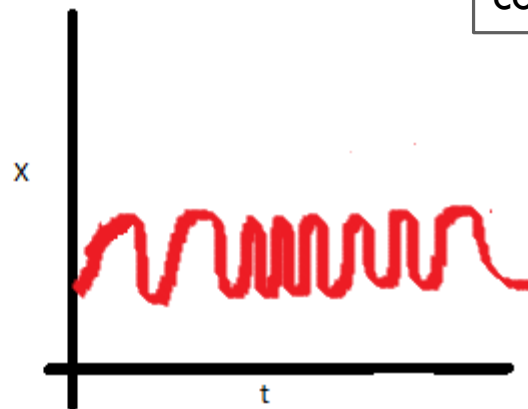
- Média do processo é constante;
- Variância do processo é constante;
- A covariância do processo não deve ser função do tempo.

A maioria das séries reais são não estacionárias.

Com métodos matemáticos é possível convertê-las em estacionárias.



Stationary series



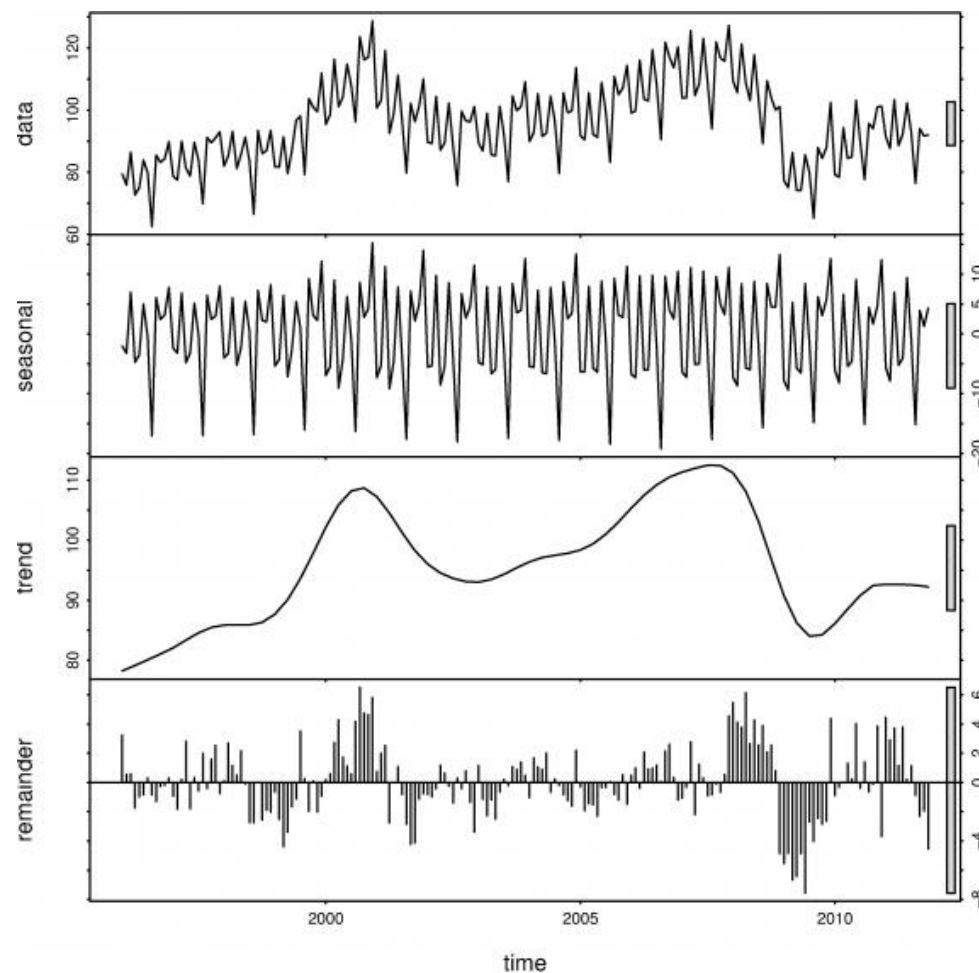
Non-Stationary series

RESUMINDO

Séries temporais são constituídas de quatro padrões:

1. **Tendência**, que é o comportamento de longo prazo da série, que pode ser causada pelo crescimento demográfico, ou mudança gradual de hábitos de consumo, ou qualquer outro aspecto que afete a variável de interesse no longo prazo.
2. **Variações cíclicas ou ciclos**, flutuações nos valores da variável, que podem ser resultado de variações da economia como períodos de crescimento ou recessão, ou fenômenos climáticos como o El Niño.
3. **Variações sazonais ou sazonalidade**, flutuações nos valores da variável, e que se repetem, geralmente em função das estações do ano (ou em função de feriados ou festas populares, etc.);
4. **Variações irregulares**, que são as flutuações inexplicáveis, resultado de fatos fortuitos e inesperados como catástrofes naturais, atentados terroristas como o de 11 de setembro de 2001, decisões intempestivas de governos, etc.

DECOMPOSIÇÃO DE UMA SÉRIE TEMPORAL



MODELOS DE PREVISÃO

- Naive
- Modelo de média móvel (MA)
- Modelo de Amortecimento Exponencial
- Modelo Auto-regressivo integrados de média móvel (ARIMA)
- Modelo Auto-regressivo Não Linear

NAIVE (INGÊNUO)

A previsão do valor da série no instante $T+1$ é apenas dado pela última observação da série em T .

Uma aplicação clássica é a previsão do preço de uma ação. (*)

MÉDIAS MÓVEIS

A previsão é a média das últimas n observações.

Um dos problemas é a definição do tamanho da janela:

- Quanto maior o valor de n , mais “suave” é a previsão;
- Se n é pequeno, a previsão tende a oscilar muito.

Característica importante: todas as observações têm o mesmo peso.

Mas na prática, as observações mais recentes tendem a ser mais relevantes.

○ que nos leva aos modelos de amortecimento exponencial.

AMORTECIMENTO EXPONENCIAL

A ideia geral é parecida com método de médias móveis, mas os pesos das observações decrescem à medida que as observações estão mais longe.

A taxa de decréscimo dos pesos é determinada por uma ou mais constantes de amortecimento. A maior dificuldade é a escolha das constantes de amortecimento.

ARMA, ARIMA E SARIMA

ARMA: processos mistos AR e MA (séries estacionárias).

Modelos mais sofisticados:

Incluem modelos não estacionários (ARIMA) e sazonais (SARIMA);

MODELOS AUTO-REGRESSIVOS NÃO LINEARES

Modelos não lineares são mais poderosos, mas precisam de mais dados de treinamento e não são tão bem comportados quanto os modelos lineares.

Exemplo: Redes Neurais.

- Pré-Processamento dos dados é importante: pode ser útil remover tendência. [[Neural network forecasting for seasonal and trend time series, 2005](#)]

ALGUMAS DEFINIÇÕES

Janela de Entrada → Quantos e quais valores passados da série devem ser utilizados;

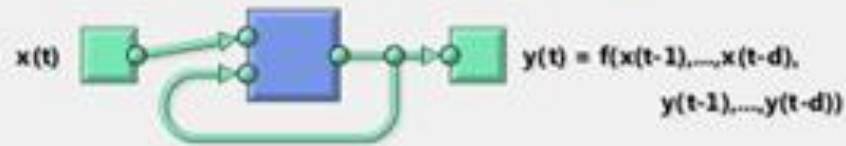
Horizonte da Previsão → A saída da rede refere-se a quantos passos à frente?

Definição de outras variáveis explicativas → Que outras variáveis podem influenciar na previsão? (exemplos: outras séries históricas (índices financeiros, temperatura), dia da semana, hora da previsão, mês da previsão, etc...)

PREVISÃO DE SÉRIES

☐ Nonlinear Autoregressive with External (Exogenous) Input (NARX)

Predict series $y(t)$ given d past values of $y(t)$ and another series $x(t)$.



☐ Nonlinear Autoregressive (NAR)

Predict series $y(t)$ given d past values of $y(t)$.



☐ Nonlinear Input-Output

Predict series $y(t)$ given d past values of series $x(t)$.

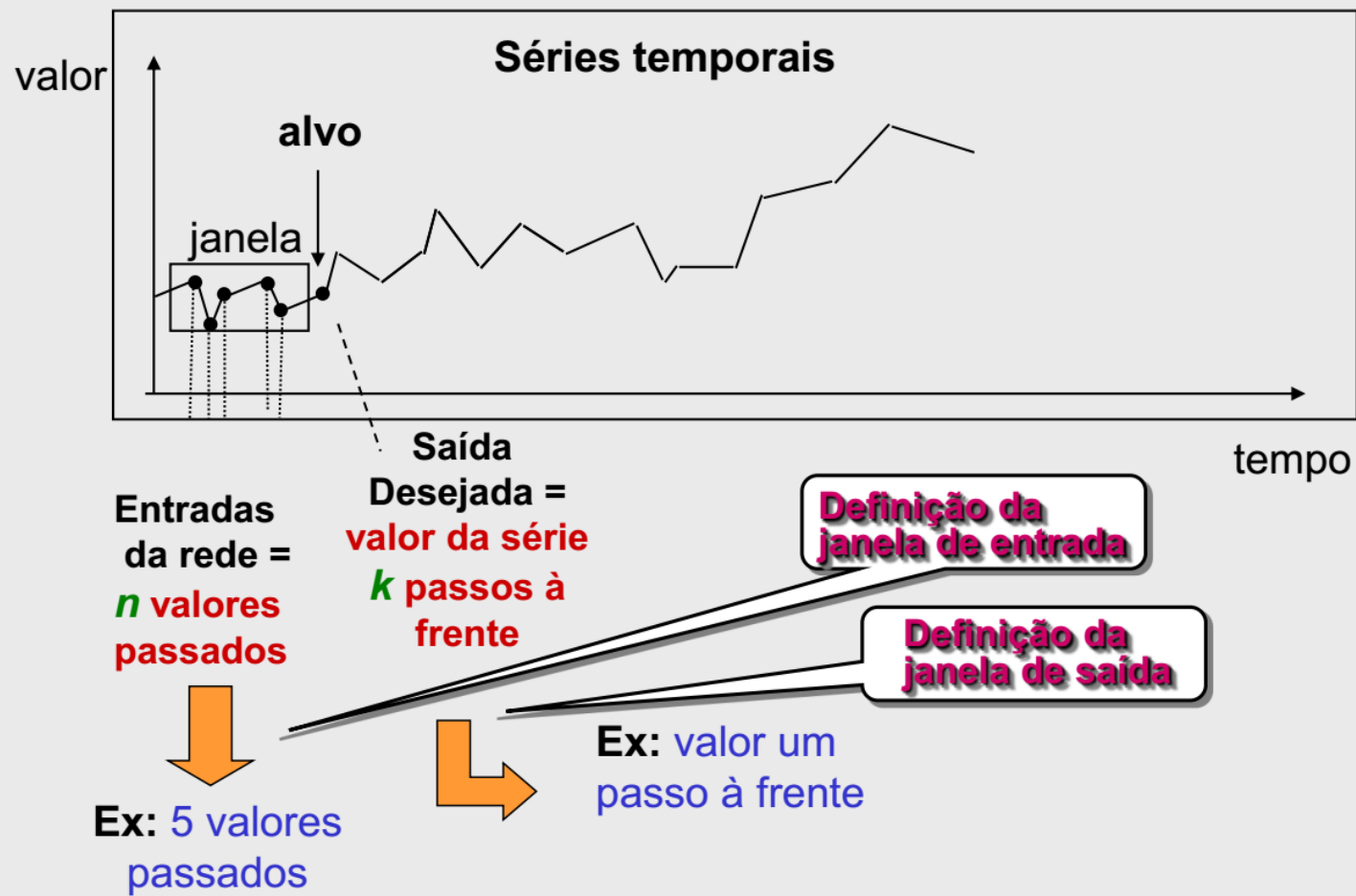
Important Note: NARX solutions are more accurate than this solution. Only use this solution if past values of $y(t)$ will not be available when deployed.



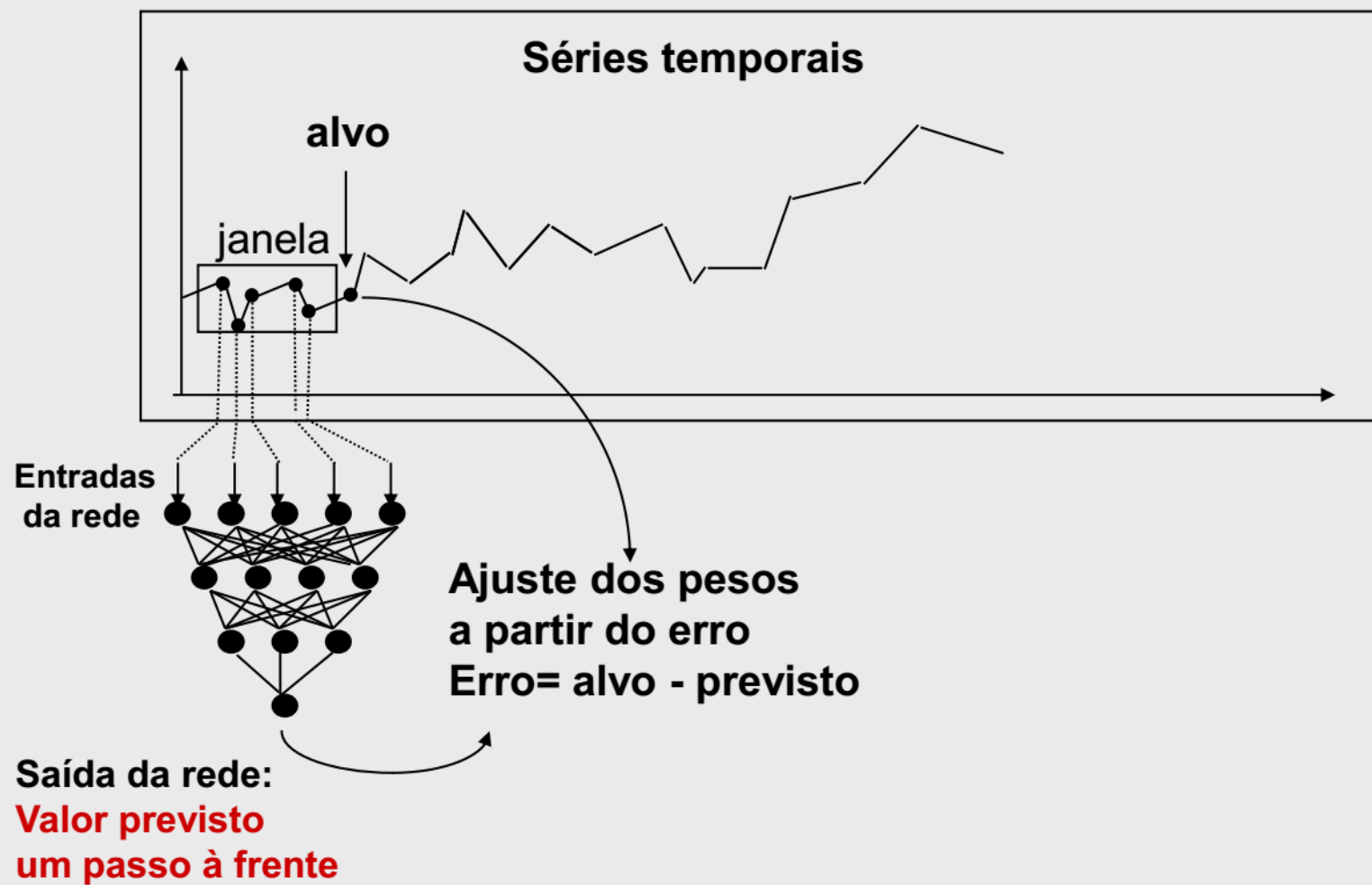
NAR – Nonlinear Autoregressive Model

Previsão de séries temporais com valores da própria série.

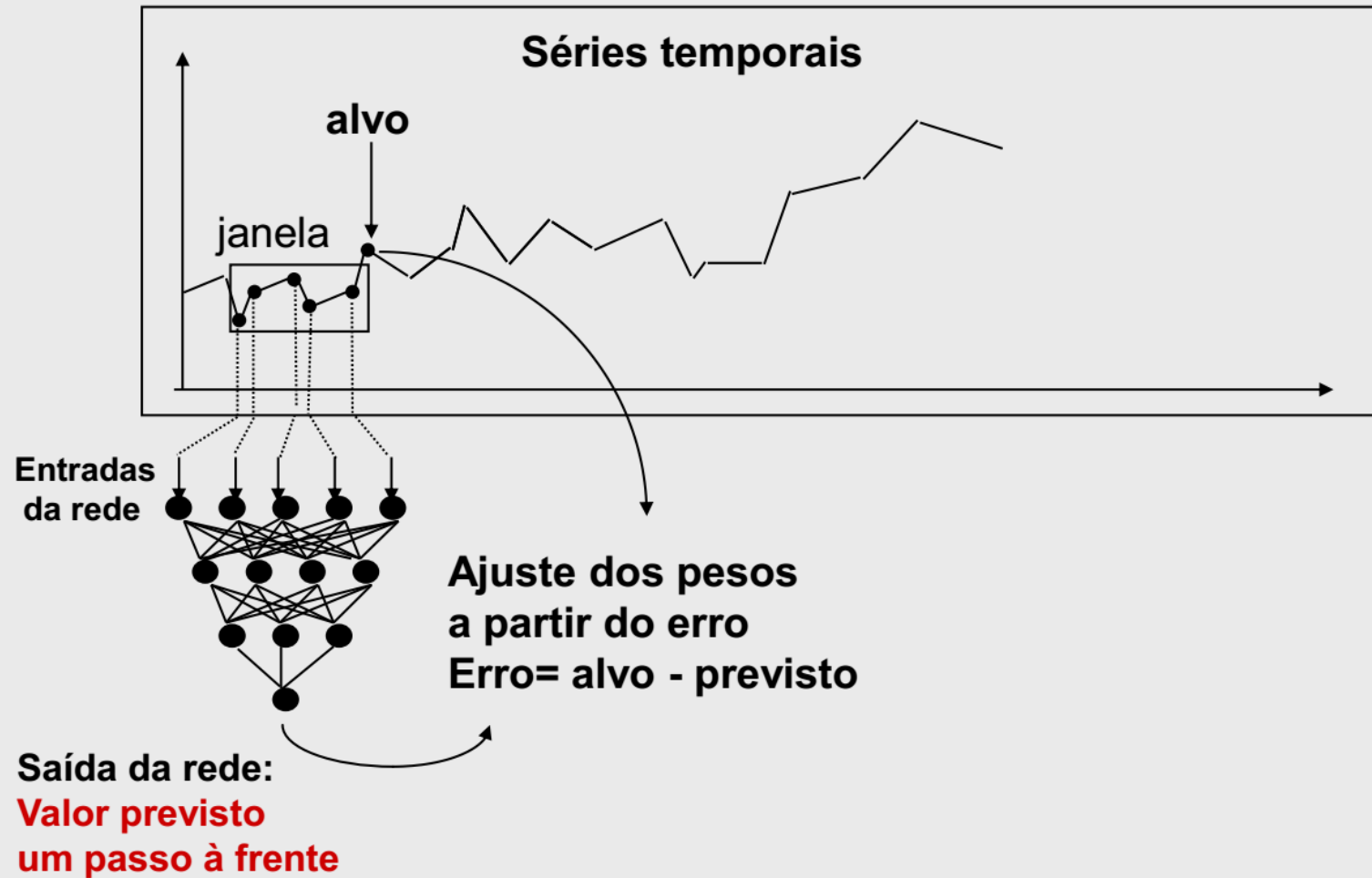
NAR



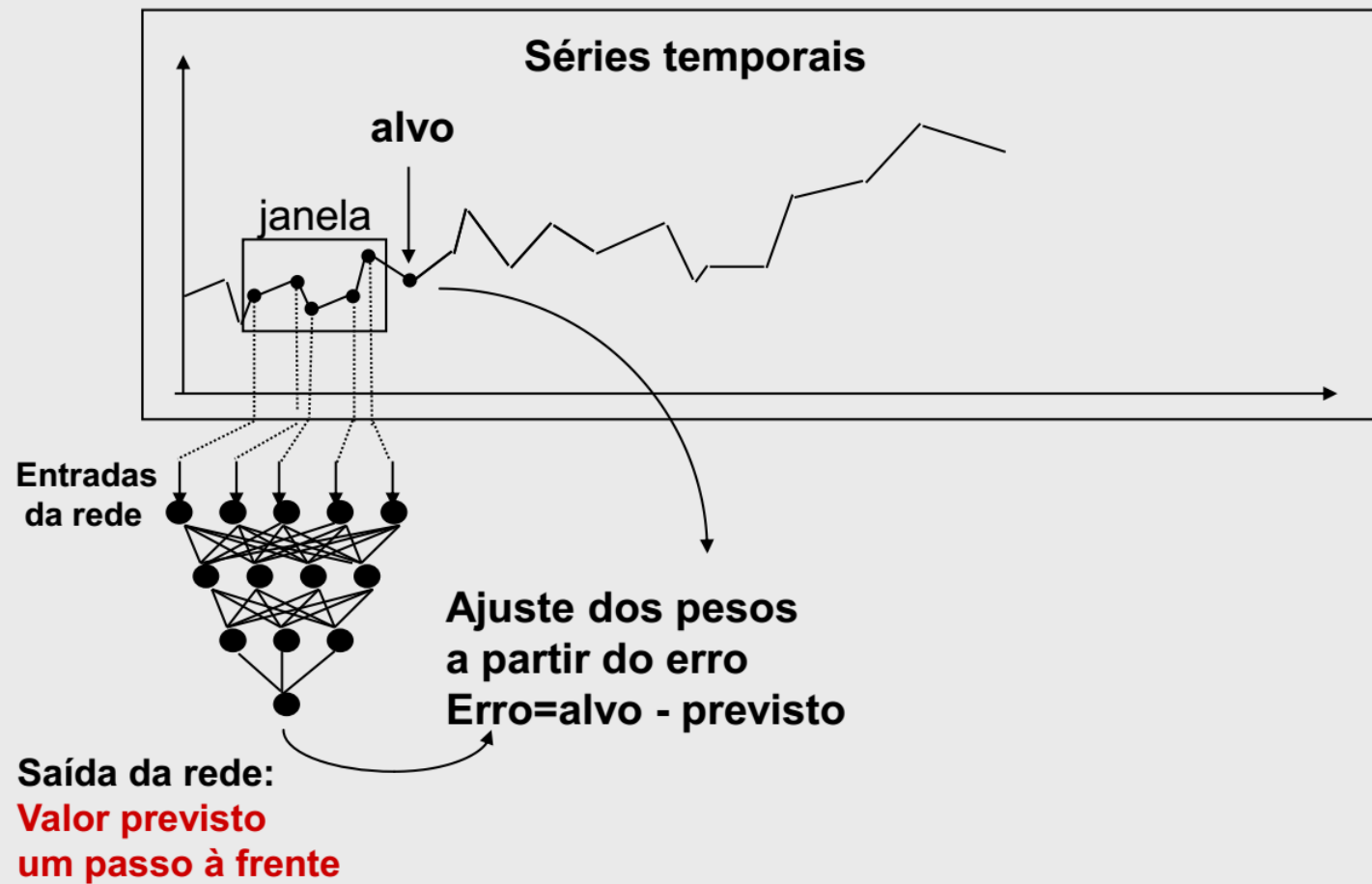
NAR



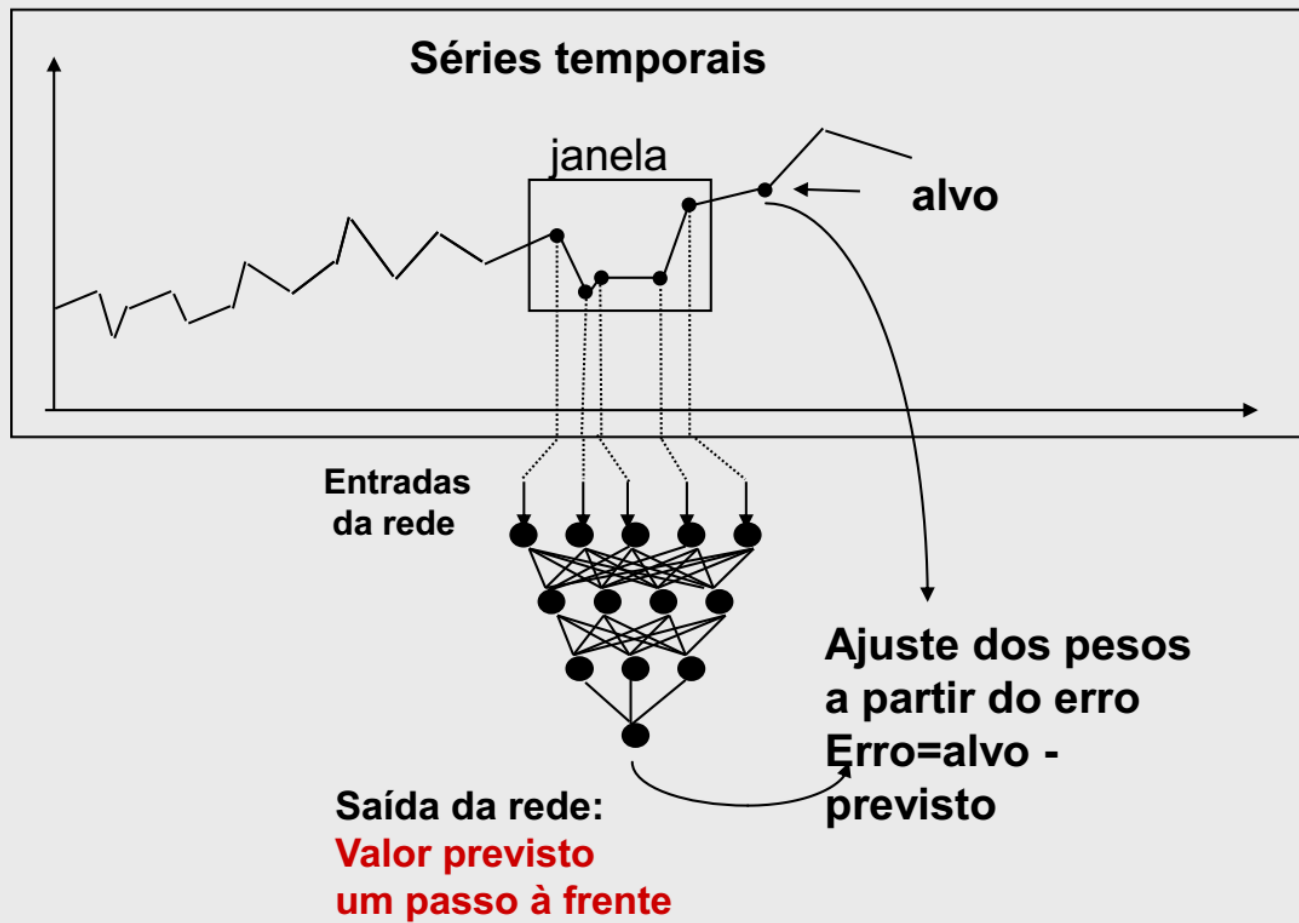
NAR



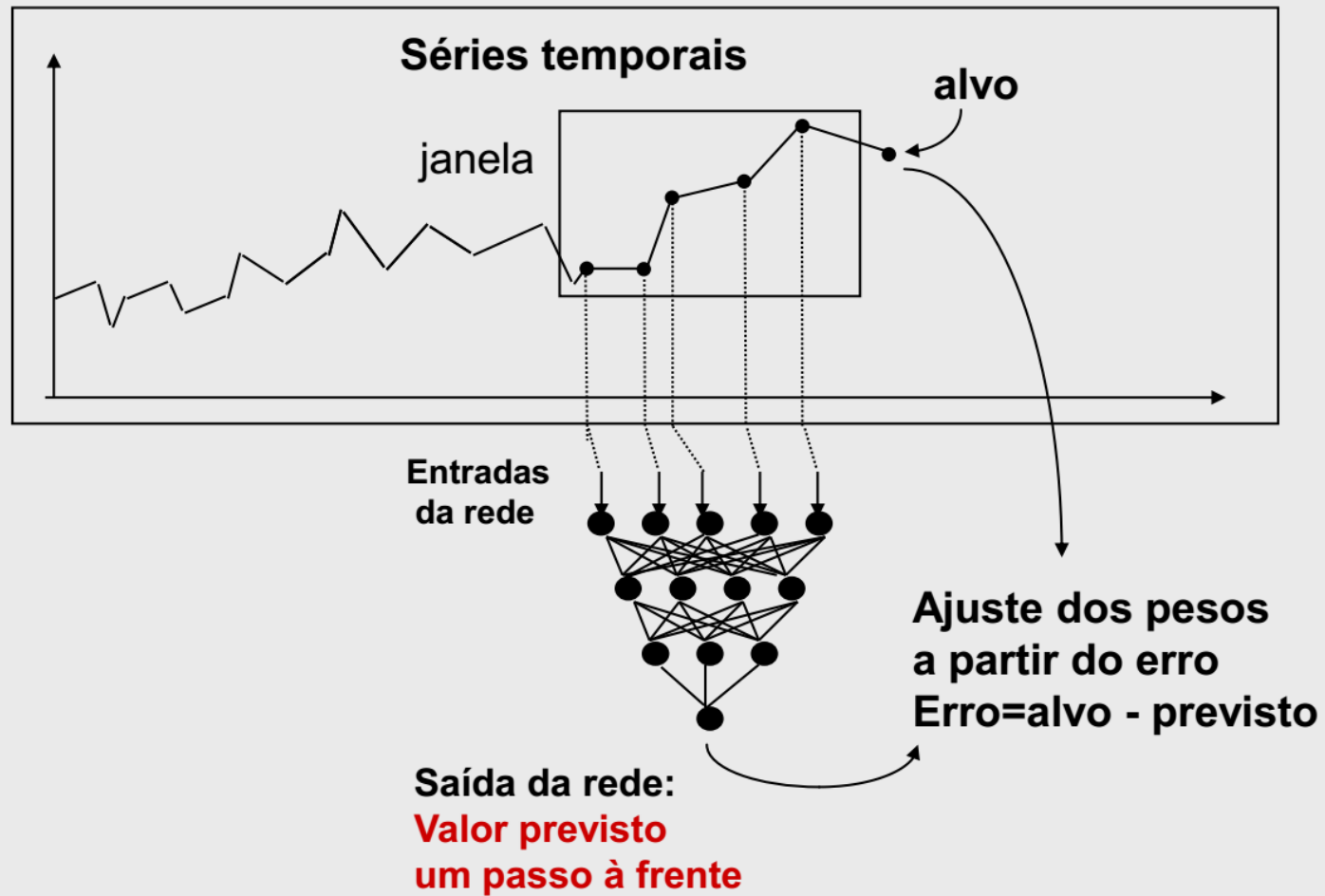
NAR



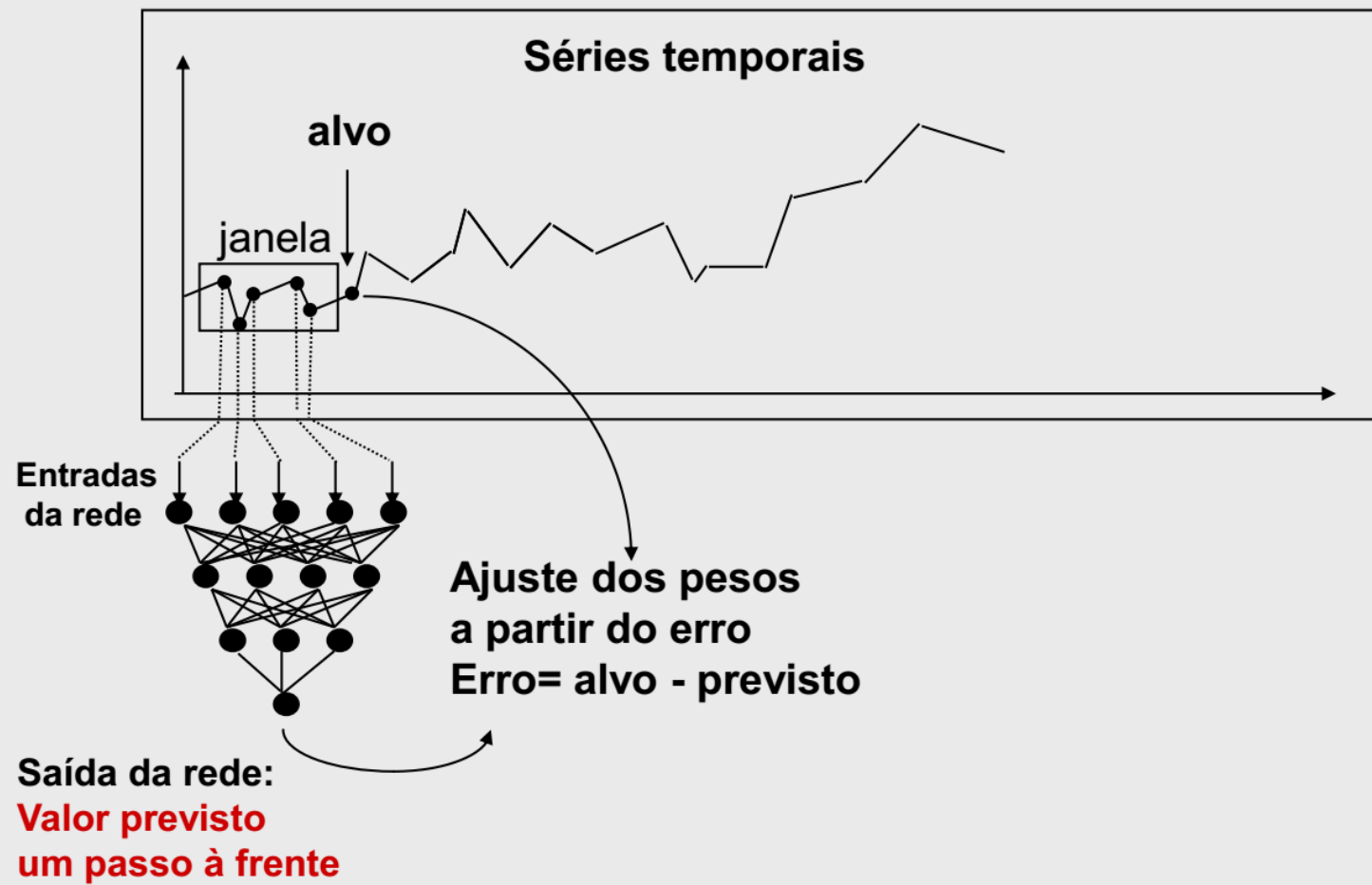
NAR



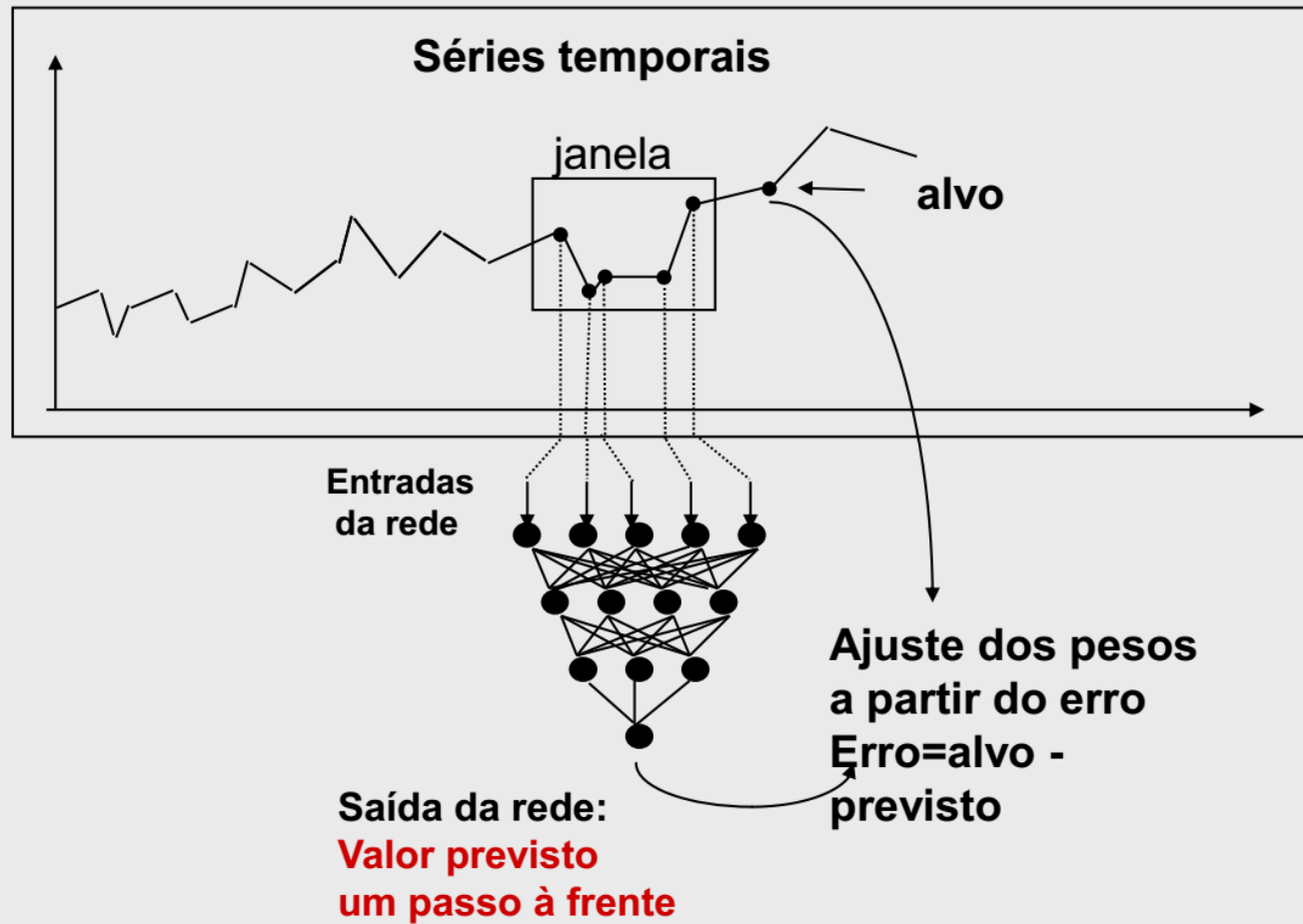
NAR



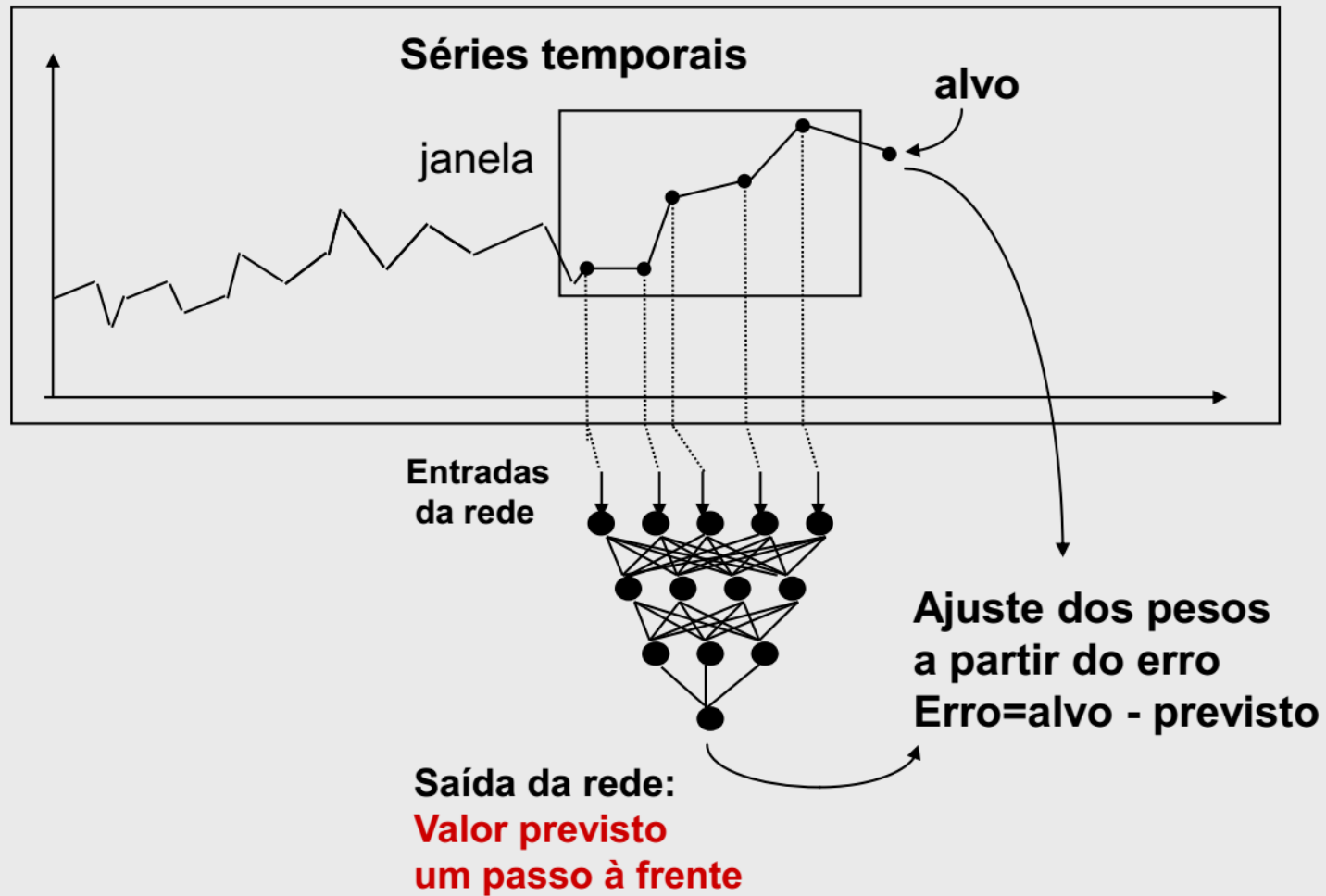
NAR



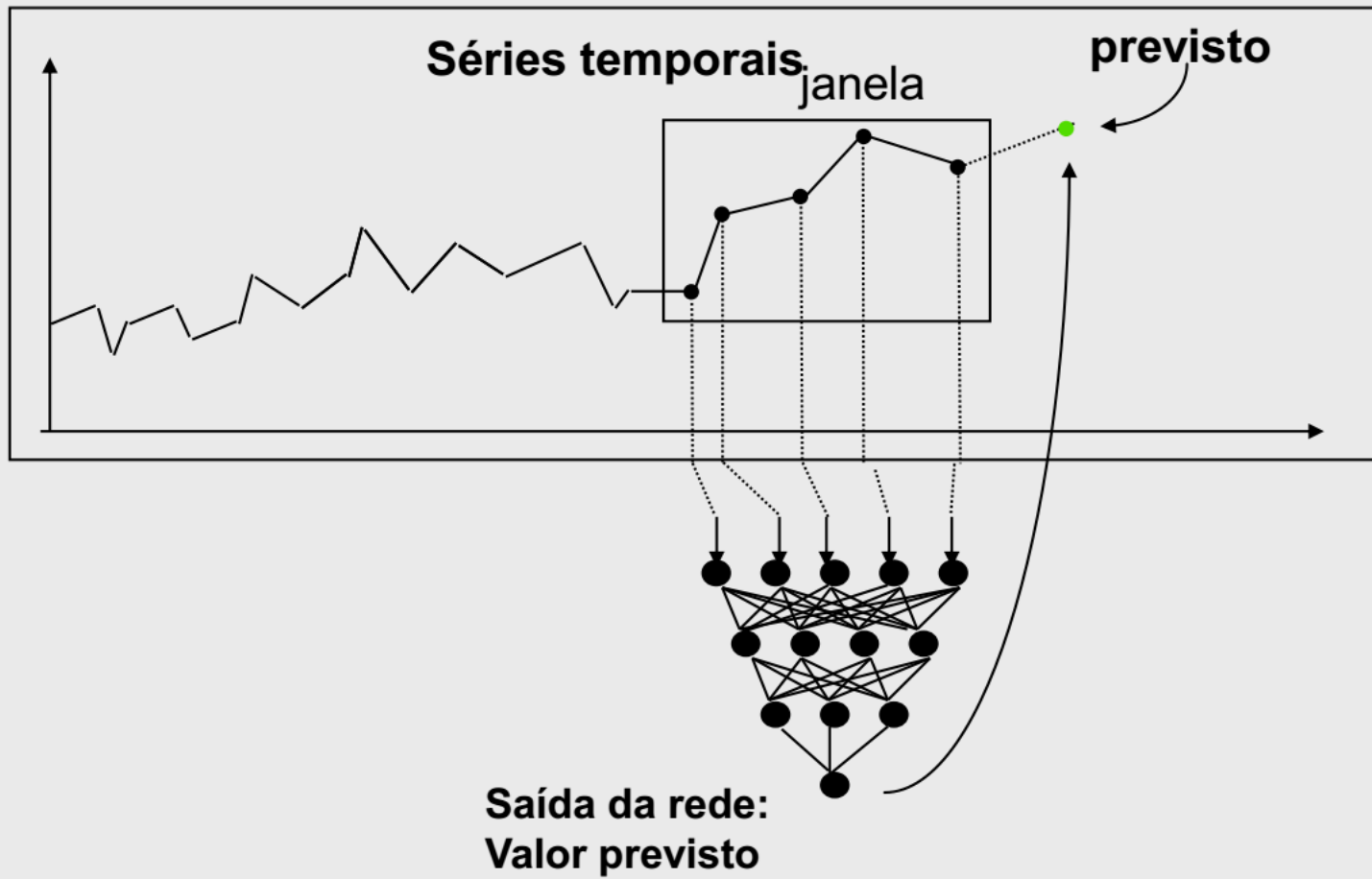
NAR



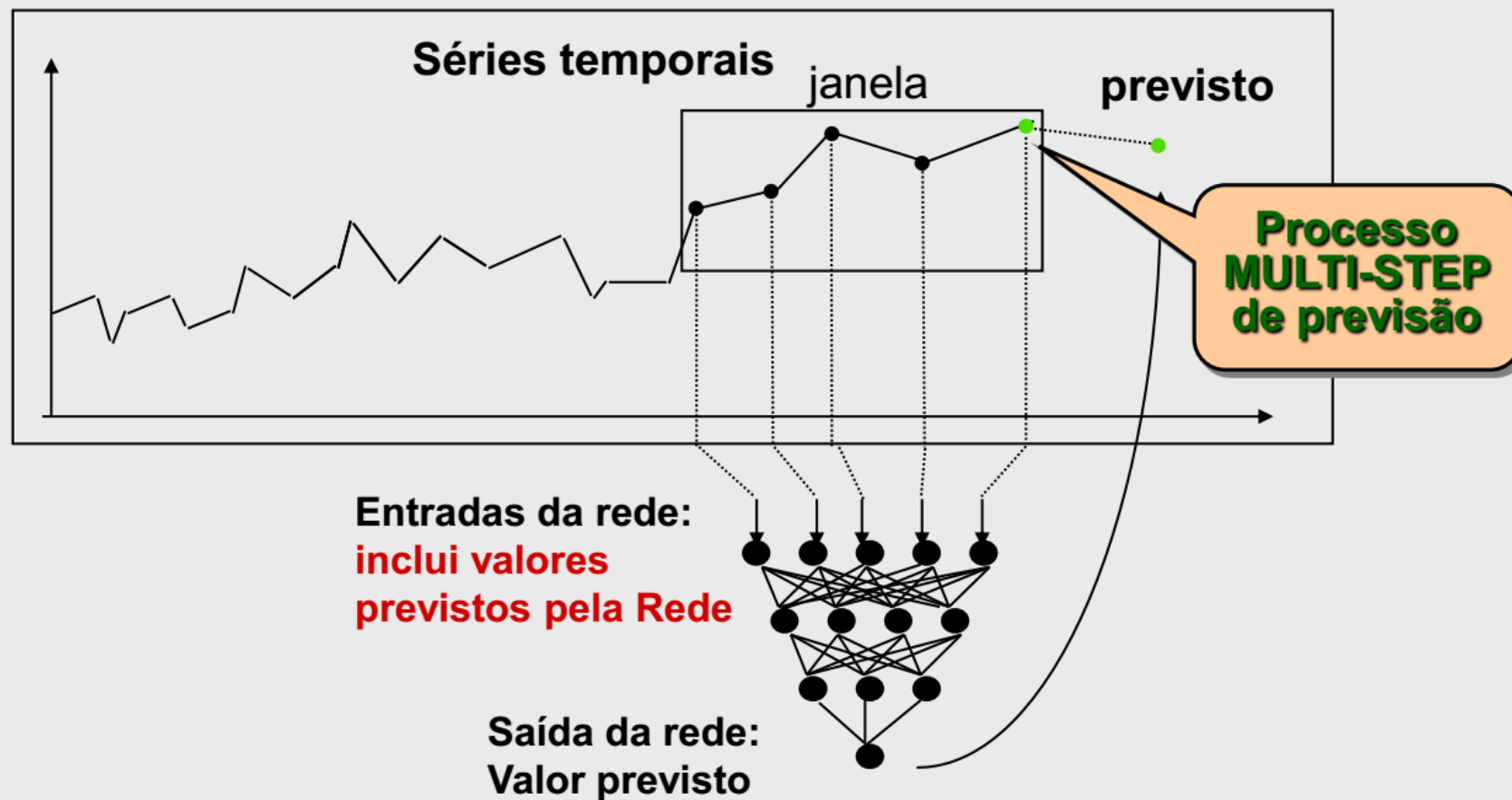
NAR



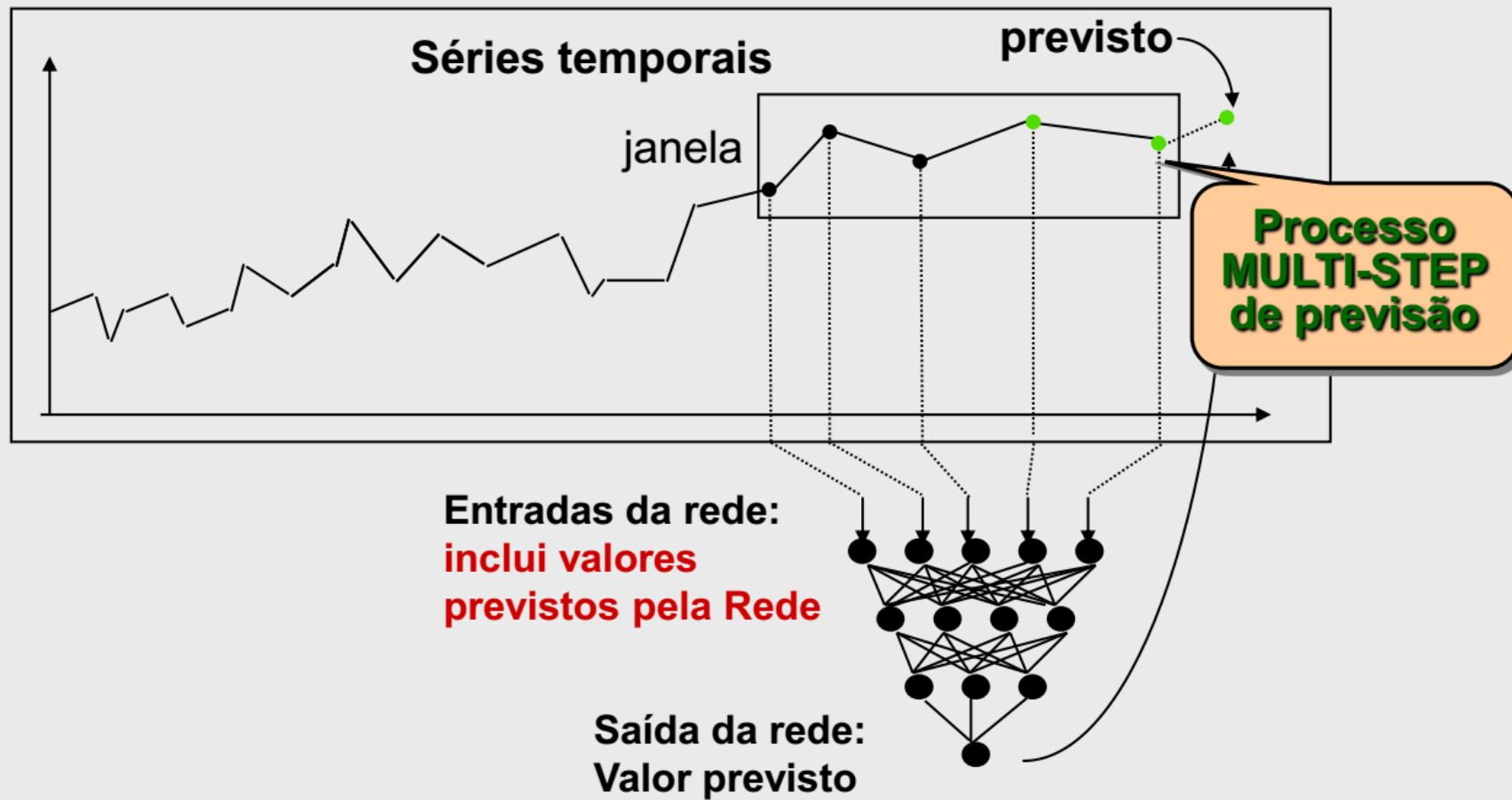
NAR



NAR



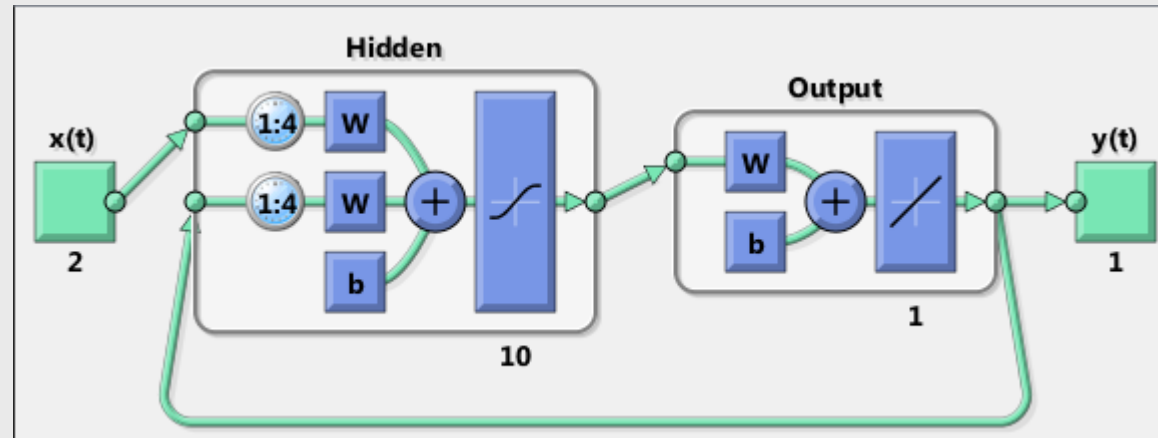
NAR



NAR

Processo Multi Step (Closed Loop):

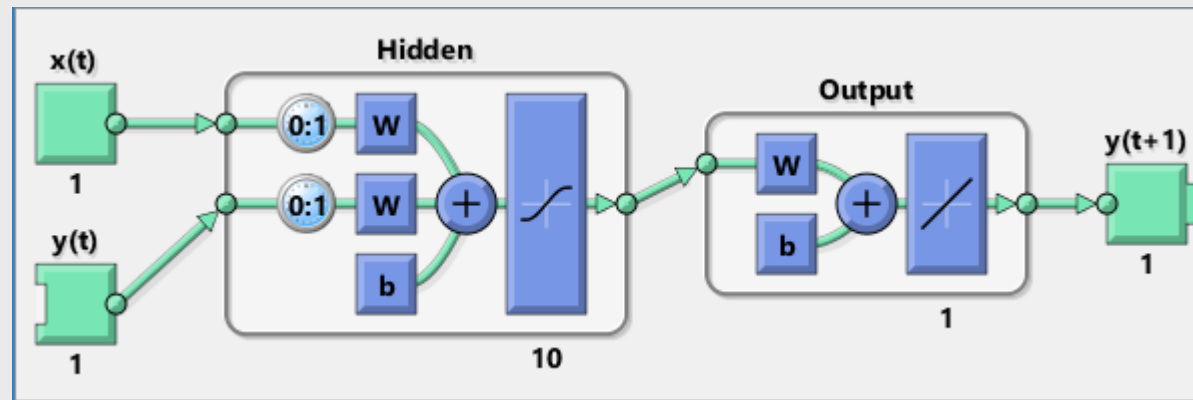
- Existe um horizonte da série antes da previsão, e deseja-se prever n passos à frente.
- A própria previsão da rede é utilizada para prever novos valores.
- Erro aumenta conforme aumenta-se o número de previsões à frente.



NAR

Processo One Step Ahead (Open Loop):

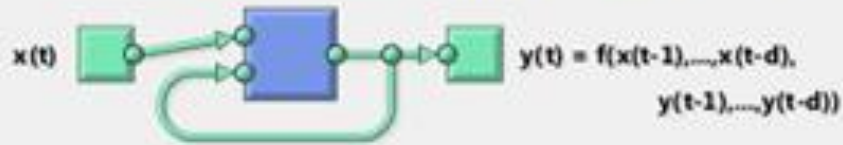
- Existe um horizonte da série antes da previsão, e deseja-se prever 1 passo à frente.
- Cada previsão somente é feita quando se tem todos os valores da série histórica imediatamente anteriores ao valor à ser previsto.
- Erro não é propagado para novas previsões.



PREVISÃO DE SÉRIES

● Nonlinear Autoregressive with External (Exogenous) Input (NARX)

Predict series $y(t)$ given d past values of $y(t)$ and another series $x(t)$.



● Nonlinear Autoregressive (NAR)

Predict series $y(t)$ given d past values of $y(t)$.



● Nonlinear Input-Output

Predict series $y(t)$ given d past values of series $x(t)$.

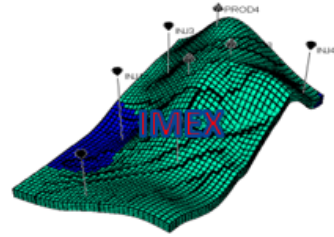
Important Note: NARX solutions are more accurate than this solution. Only use this solution if past values of $y(t)$ will not be available when deployed.



NARX – Nonlinear Autoregressive Model with External

Previsão de séries temporais com valores da própria série e valores de outra série.

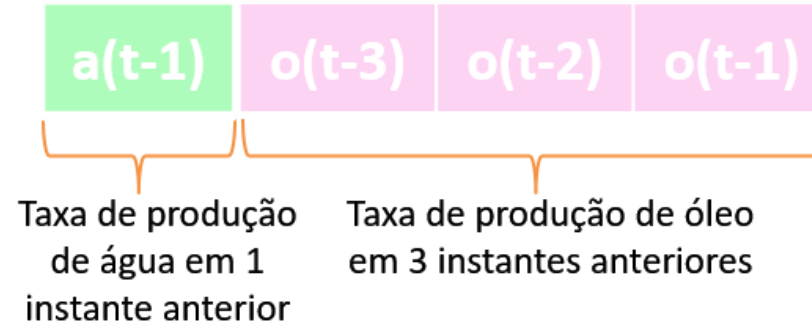
NARX



1 simulação



Horizonte: 5 anos
Janela óleo: 3 meses
Janela água: 1 mês



Saída da Rede

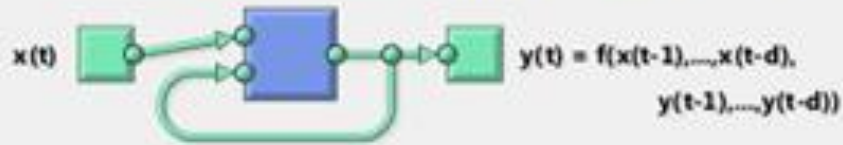


	Janela água	Janela óleo			
11/2011	A(3)	O(1)	O(2)	O(3)	O(4)
12/2011	A(4)	O(2)	O(3)	O(4)	O(5)
01/2012	A(5)	O(3)	O(4)	O(5)	O(6)
•					
•					
•					
08/2016	A(59)	O(57)	O(58)	O(59)	O(60)

PREVISÃO DE SÉRIES

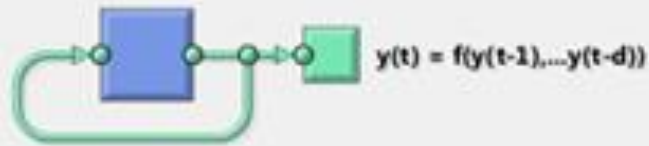
● Nonlinear Autoregressive with External (Exogenous) Input (NARX)

Predict series $y(t)$ given d past values of $y(t)$ and another series $x(t)$.



● Nonlinear Autoregressive (NAR)

Predict series $y(t)$ given d past values of $y(t)$.



● Nonlinear Input-Output

Predict series $y(t)$ given d past values of series $x(t)$.

Important Note: NARX solutions are more accurate than this solution. Only use this solution if past values of $y(t)$ will not be available when deployed.



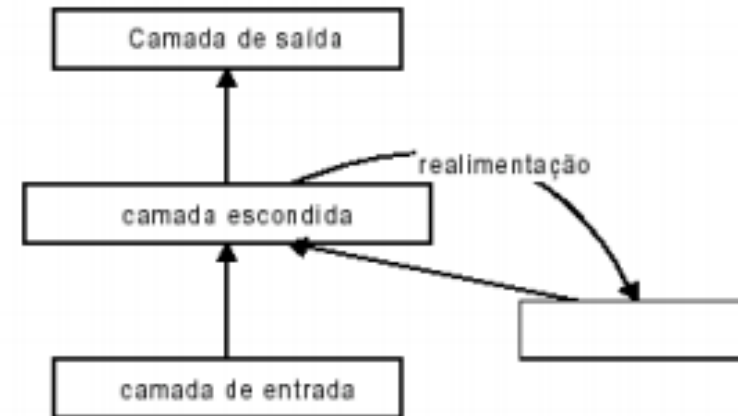
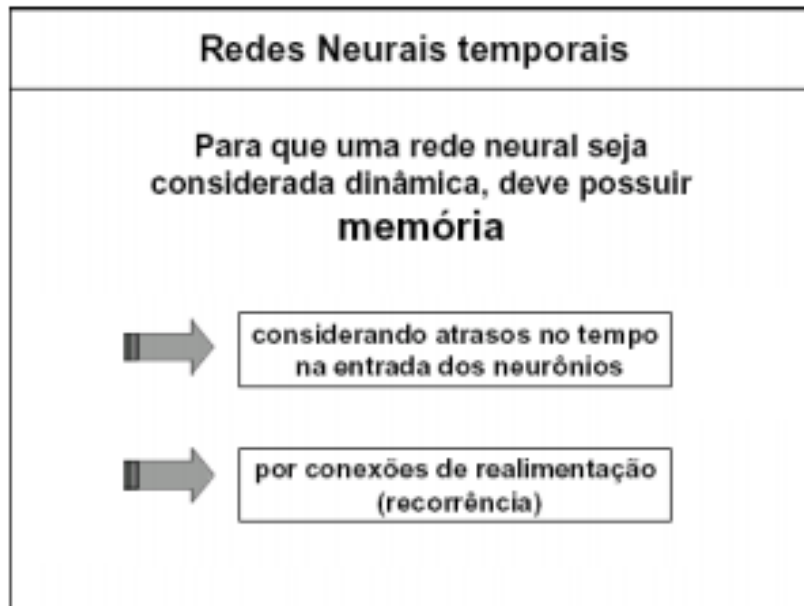
Nonlinear Input-Output

Previsão de séries temporais com valores de outra série.

Atenção: Soluções via NARX são mais precisas que essa solução.

Só utilize esta opção quando não estiverem disponível os dados da própria série a ser prevista.

REDE RECORRENTE



ESTUDO DE CASO

Previsão de Carga Elétrica

Um sistema preciso de previsão de carga oferece:

**segurança,
confiabilidade e
economia**

na operação de sistemas de potência.

PREVISÃO DE CARGA

Previsão de Curto/Curtíssimo Prazo:

- assegura uma operação eficiente e segura do sistema.

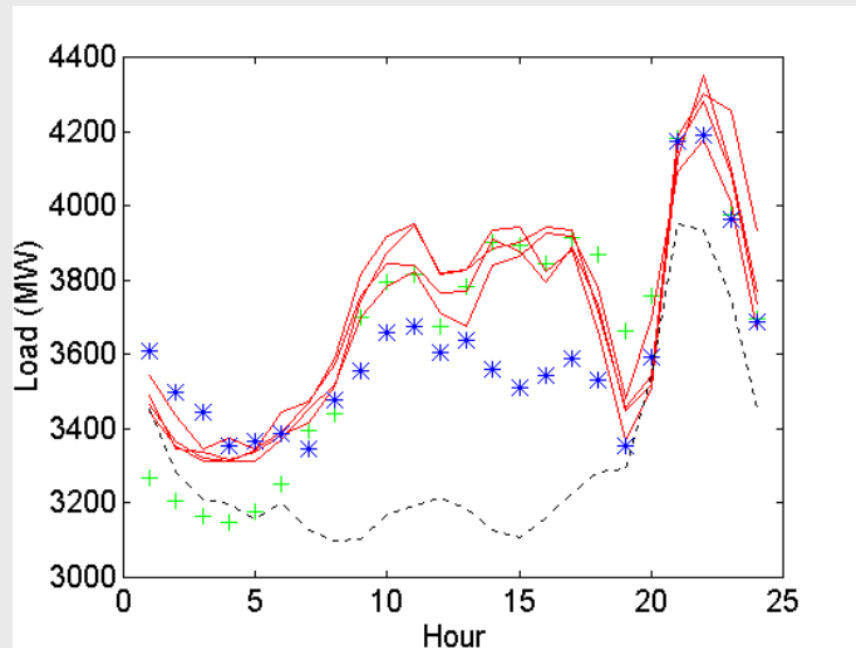
Previsão de Longo Prazo

- expansão e planejamento da operação.

Previsão de Carga Horária

- Previsão da carga 168 e 24 horas a frente (CEMIG)

PERFIL DE CARGA HORÁRIA



- - Terça, Quarta, Quinta e Sexta-Feira
- ****** - Sábado
-** - Domingo
- +++** - Segunda-Feira

PREVISÃO DE CARGA HORÁRIA

4 redes diferentes de acordo com o dia da semana

Entradas:

- janela de 5 valores passados
- valor da carga 7 dias antes, no mesmo horário
- codificação binária da hora a ser prevista

Treinamento:

- Conjunto de Treinamento: 2 meses anteriores
- Re-treinamento a cada mês

PREVISÃO DE CARGA

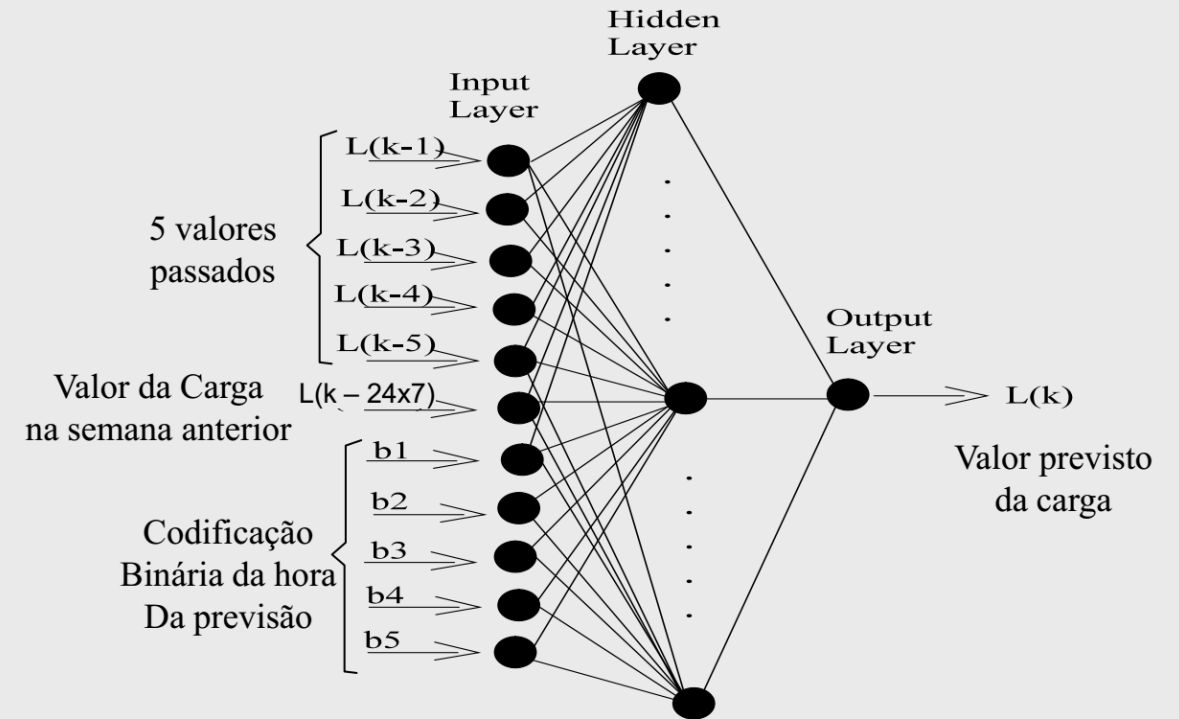
Topologia da Rede

Camada de entrada: 11 neurônios.

- 5 últimos valores de carga.
- Carga no mesmo dia e hora na semana anterior.
- Codificação binária do horário.

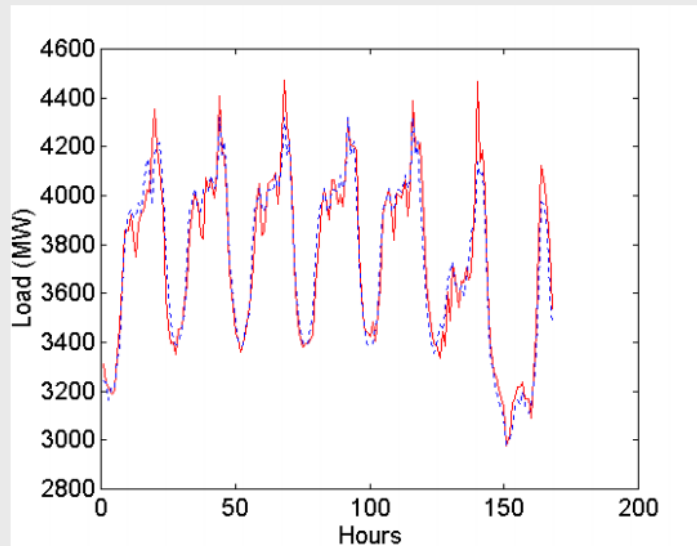
Camada escondida: 20 neurônios.

Camada de saída: 1 neurônio.



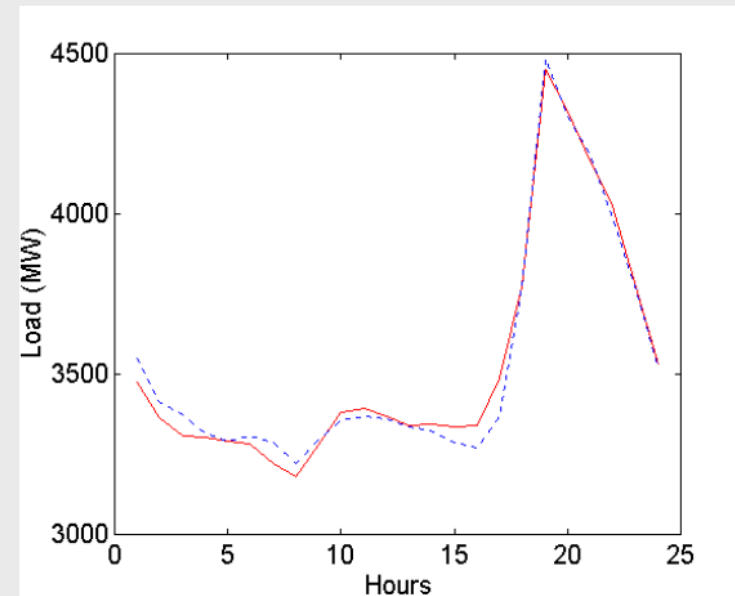
PREVISÃO DE CARGA

**Uma semana em Março
168 horas a frente**



— **Valores Reais**
- - - **Valores Previstos**

**1º de agosto
24 horas a frente**



— **Valores Reais**
- - - **Valores Previstos**

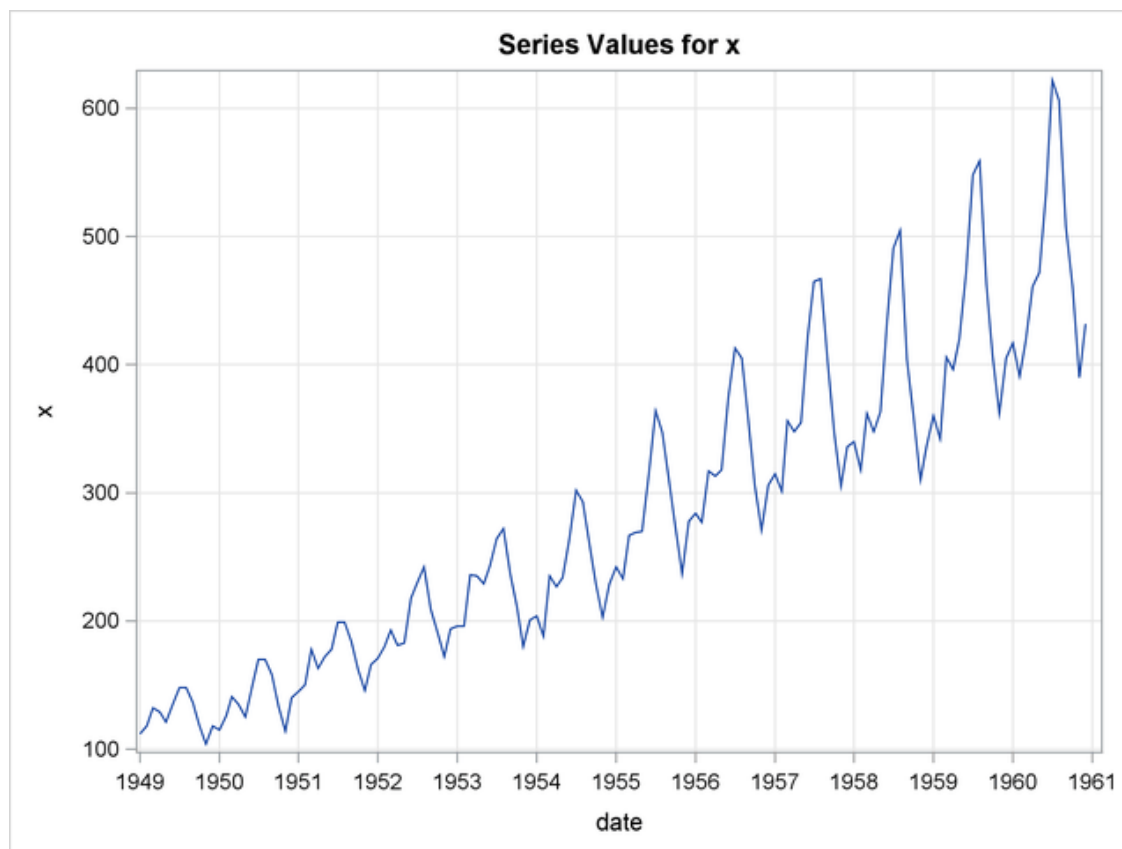
ESTUDO DE CASO

Faturamento Varejo



- Mês
- Ano
- Faturamento

ESTUDO DE CASO



**Passageiros de
companhias aéreas.**