

Análise de Séries Temporais com Deep Learning

Advanced Institute for Artificial Intelligence – AI2

<https://advancedinstitute.ai>

Agenda

- ☐ Séries Temporais
- ☐ Conceitos
- ☐ Ferramentas para Manipular Datas
- ☐ Exploracao Estatística
- ☐ Conceitos

Séries Temporais:

- ☐ Conjunto de observações realizadas sequencialmente ao longo do tempo.
- ☐ Muitos problemas são modelados usando séries temporais
 - Finanças, marketing, ciências sociais, entre outras
- ☐ informações históricas, em muitos casos, auxiliam o entendimento e permite previsões

Alguns conceitos relacionados com Séries Temporais:

- Séries Univariadas:

- Um único valor ao longo do tempo

- Séries Multivariadas:

- Dois ou mais valores variam ao longo do tempo

Análise de Séries Temporais com Deep Learning

Tempo	Característica 1	Característica 2	
1	10	10	}
2	20	20	
1	5	34	}
2	6	55	
1	3	60	}
2	7	90	

Passo temporal

Passo temporal

Passo temporal

Característica 1	Característica 2	Característica 1	Característica 2
10	10	20	20
5	34	6	55
3	60	7	90

Análise de Séries Temporais com Deep Learning

Tempo	Característica 1	Característica 2
1	10	10
2	20	20
3	5	34
4	6	55
5	3	60
6	7	90
1	5	5
2	8	9

Característica 1	Característica 2	Característica 1 (t+1)	Característica 2 (t+2)
10	10	20	20
20	20	5	34
5	34	6	55
6	55	3	60
3	60	7	90
5	5	8	9

Quebra da série
para outra
amostra

Alguns conceitos relacionados com Séries Temporais:

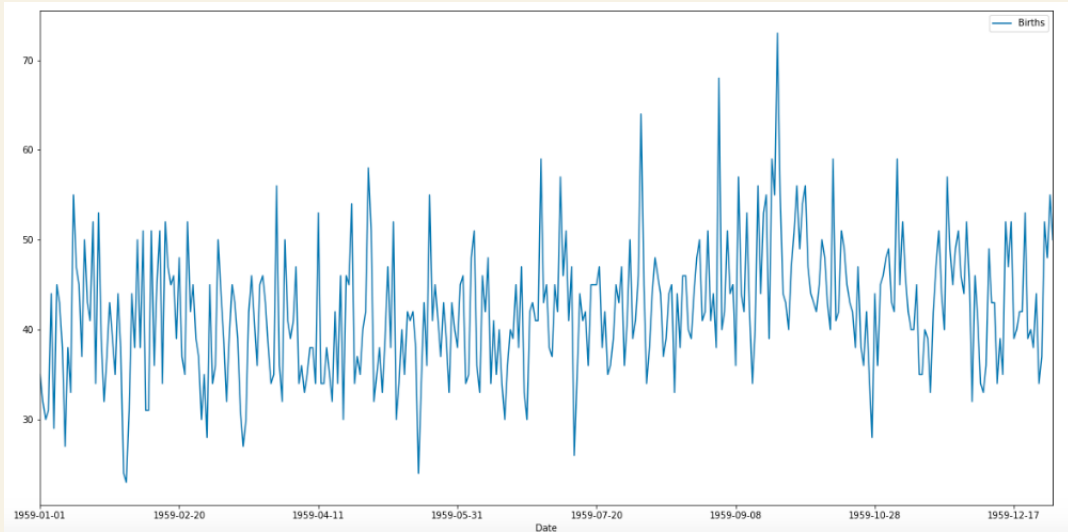
- ☐ Séries Estacionárias

- As características apresentam padrões que não variam com o tempo

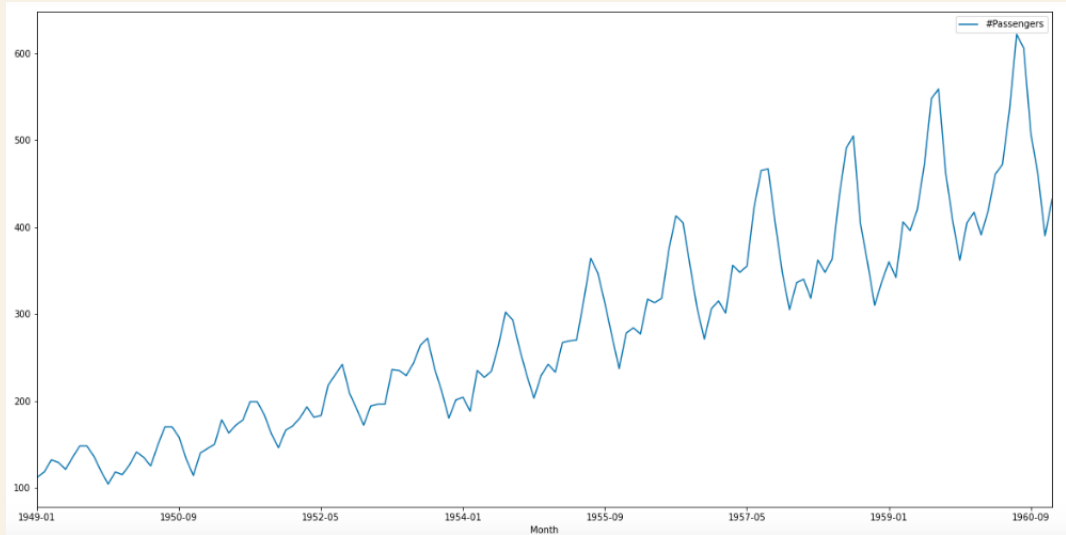
- ☐ Séries não-Estacionárias

- Padrões mudam de modo sistemático com dependência em relação ao tempo

Análise de Séries Temporais com Deep Learning



Análise de Séries Temporais com Deep Learning



Avaliando se uma série temporal é estacionária. Métodos possíveis:

- ☐ Analisar gráficos: você pode revisar um gráfico de série temporal de seus dados e verificar visualmente se há tendências ou sazonalidades óbvias
- ☐ Estatísticas resumidas: você pode revisar as estatísticas resumidas dos seus dados para temporadas ou partições aleatórias e verificar se há diferenças óbvias ou significativas
- ☐ Testes estatísticos: você pode usar testes estatísticos para verificar se as expectativas de estacionariedade foram atendidas ou foram violadas.

Séries temporais não-estacionárias podem conter as seguintes características:

- ☐ Tendência: um componente linear sistemático geral ou (na maioria das vezes) componente não linear que muda com o tempo e não se repete
- ☐ Sazonalidade: ciclos que se repetem ao longo do tempo
- ☐ Ruídos : um componente não sistemático que não é Tendência / Sazonalidade nos dados

Predição:

- ☐ histórico da série sendo usada para aprender o comportamento da série, permitindo prever valores para os períodos futuros
- ☐ Muitos métodos estatísticos exigem que as séries sejam estacionárias para realizar tais previsões
 - ☐ Nesse cenário alguns métodos podem ser usados para eliminar, tendências, sazonalidades e ruídos
- ☐ Métodos de Deep Learning podem fazer explorações mais amplas nesse cenário

- Aprendizagem de máquina são aplicados normalmente para modelar problema utilizando recursos além dos métodos estatísticos tradicionais
- É difícil prever se um algoritmo oferecerá melhor desempenho para um problema considerando que a série é estacionária ou não
- Nesse sentido, a fonte de informação ainda que não seja uma série estacionária, pode ser usada na engenharia e seleção de features

- Segmentação dos dados para treino, validação e testes
- Normalmente os dados são divididos escolhendo uma posição específica do dataset, ou aleatoriamente
- no caso de séries temporais são escolhidos períodos
 - Período de treino
 - Período de validação
 - Período de teste

- ❑ Objeto datetime é disponibilizado pelo python para manipular datas
- ❑ Numpy e Pandas oferecem recursos para manipular datas
 - Slicing
 - Mudança de frequência
 - Segmentação
 - Agrupamento

Representação de série para Deep Learning:

- A sequência é dividida em etapas fixas
 - Uma quantidade de dias, horas, semanas, elementos, etc
 - Para cada elemento da sequência um valor é associado, que é o valor da sequência
 - Esse é o conjunto de features
- O próximo item da sequência é a coluna alvo

RNN são redes neurais estruturadas de forma a representar sequências

- Para isso tais redes são montadas com base em um tipo de camada chamada camada recorrentes
- Tais camadas possuem uma unidade interna chamada memória
 - A idéia dessa unidade é processar sequências de quaisquer tamanhos

Camada recorrente:

- ☐ Camadas recorrentes recebem como entrada uma matrix multidimensional
- ☐ Cada item da sequência gera um valor de saída y e armazena uma matriz com um valor intermediário na unidade de memória
- ☐ O próximo valor é calculado usando o valor x da sequência e o valor intermediário gerado no valor anterior
- ☐ Esse laço é repetido ao longo de todos os valores da sequência

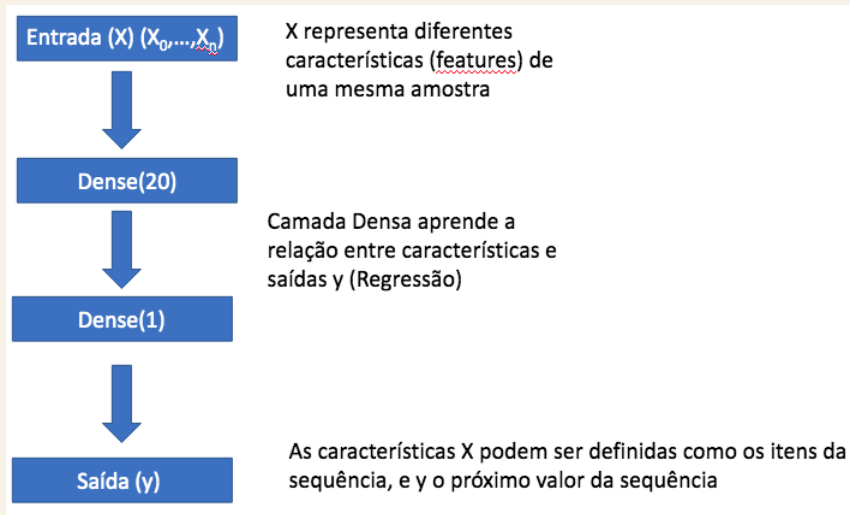
Dimensionalidade de uma camada recorrente:

- ☐ Tamanho do batch
- ☐ Passos temporais
- ☐ Dimensionalidades
 - Para séries univariadas esse valor é 1
 - Para séries com mais valores esse valor é maior

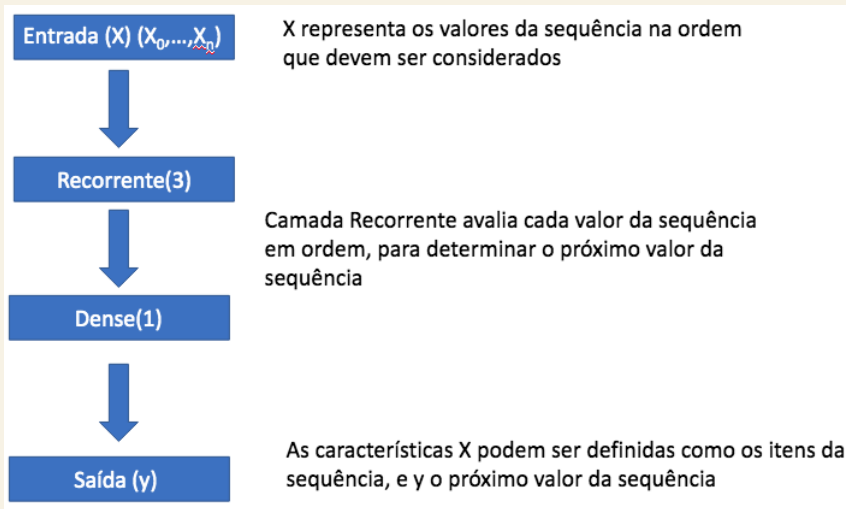
Uma rede MLP pode ser usada para prever o próximo valor em uma sequência:

- ☐ Todos os elementos da sequência são considerados como uma característica unificada do processo
- ☐ Para que a MLP possa identificar tendências, sazonalidades
- ☐ Dimensionalidades

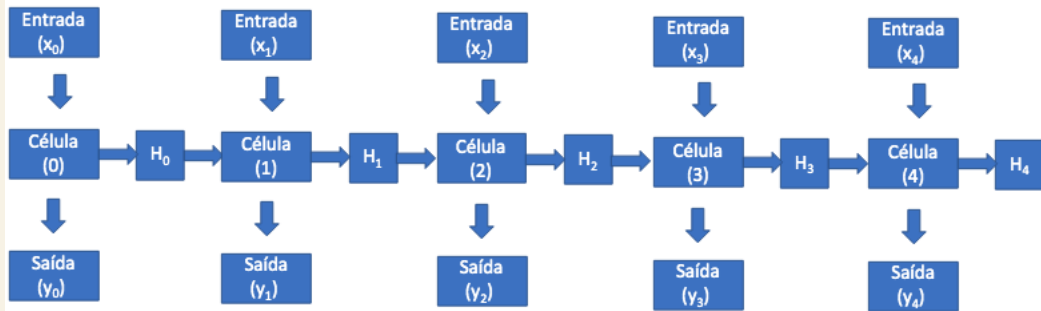
Análise de Séries Temporais com Deep Learning



Análise de Séries Temporais com Deep Learning



Sequência de Tamanho 5



- As entradas são tridimensionais. Por exemplo, se tivermos um tamanho de janela de 30 registros de data e hora e os agruparmos em tamanhos de quatro, a forma será $4 \times 30 \times 1$
- A cada registro de data e hora, a entrada da célula de memória será uma matriz 4×1
- A célula também receberá a entrada da matriz de estados da etapa anterior.
 - No primeiro passo será zero.
 - Para os seguintes, será a saída da célula de memória.
- Além do vetor de estado, é claro que a célula produzirá um valor Y

- Se a célula de memória é composta por três neurônios, a matriz de saída será 4×2 (tamanho de batch=4 neuronios=3)
- Portanto, a saída total da camada é tridimensional, neste caso, $4 \times 30 \times 3$.
- Com quatro sendo o tamanho do lote, três sendo o número de unidades e 30 sendo o número de etapas gerais.

- Em uma RNN simples, a saída de estado H é apenas uma cópia da matriz de saída Y .
- Por exemplo, H_0 é uma cópia de Y_0 , H_1 é uma cópia de Y_1 e assim por diante.
- Portanto, a cada registro de data e hora, a célula de memória obtém a entrada atual e também a saída anterior.

- ❑ Agora, em alguns casos, você pode inserir uma sequência, mas não deseja produzir e deseja apenas um único vetor para cada instância do lote. Isso geralmente é chamado de `sequence to vector`.
- ❑ Para isso basta ignorar todas as saídas, exceto a última.
- ❑ Esse comportamento é padrão do `eras` é definido pela variável `return_sequences=false`
- ❑ Se deseja retornar a sequência inteira basta definir `return_sequences=true`
- ❑ Isso será necessário quando utilizar diversas camadas recorrentes empilhadas

Algoritmos de predição utilizam amostras conhecidas do problema para treino (Período de treino)

- Quando temos apenas uma única sequência de eventos é difícil identificar os dados que generalizam o problema
 - Nesse caso, é mais interessante separar segmentos da série como amostras
- Objeto dataset do keras permite montar amostras de séries a partir de uma sequência unica

- ☐ Objeto Dataset Keras
- ☐ Obtendo dados do numpy
- ☐ Obtendo dados de um CSV

Análise de Séries Temporais com Deep Learning

Fontes:

- ❑ [1] https://pandas.pydata.org/pandas-docs/stable/user_guide/timeseries.html
- ❑ [2] <https://jakevdp.github.io/PythonDataScienceHandbook/03.11-working-with-time-series.html>
- ❑ [3] <https://www.coursera.org/learn/tensorflow-sequences-time-series-and-prediction>
- ❑ [4] <https://data.mendeley.com/datasets/n8827ssnv7/2>
- ❑ [5] <http://plenar.io/>
- ❑ [6] <https://datascience.stackexchange.com/questions/997/where-can-i-find-free-spatio-temporal-dataset-for-download>
- ❑ [7] <https://opendata.stackexchange.com/questions/13749/looking-for-spatio-temporal-datasets>
- ❑ [8] <https://machinelearningmastery.com/time-series-data-stationary-python/>