

EA072 – Inteligência Artificial em Aplicações Industriais (2s2020) Atividades Computacionais Extras Não são consideradas no processo de avaliação

Questão 1)

Dentre outras opções na literatura, nos seguintes links:

- <https://medium.com/dair-ai/building-rnns-is-fun-with-pytorch-and-google-colab-3903ea9a3a79>
- <https://github.com/PacktPublishing/Advanced-Deep-Learning-with-Keras/blob/master/chapter1-keras-quick-tour/rnn-mnist-1.5.1.py>
- <https://github.com/zonghua94/mnist>
- https://github.com/cjblackout/LSTM_MNIST

é possível encontrar implementações que abordam o problema de classificação dos dígitos da base MNIST empregando redes neurais recorrentes, nem todos utilizando o framework Keras. Para tanto, a imagem é varrida por linha e/ou por coluna, visando converter uma informação espacial em uma sequência temporal. Procure se familiarizar com os códigos envolvidos, execute o treinamento e compare o desempenho com aqueles produzidos pelos classificadores linear, ELM, MLP e CNN.

Questão 2)

São 3 partes envolvidas nesta questão. Todas elas vinculadas à predição de séries temporais. Para mais detalhes acerca do assunto e das etapas de pré-processamento envolvidas, favor consultar o apêndice deste roteiro de atividades.

Parte 1: Execute o notebook **Q2P1_NYSP.ipynb**, constatando as funcionalidades envolvidas em cada trecho de código e analisando os resultados envolvidos. Atividades práticas:

- 2.1.1. Forneça o notebook devidamente executado, contendo, ao final, um resumo geral do que foi trabalhado em cada seção.
- 2.1.2. Com o foco no desempenho do preditor linear, procure apresentar argumentos capazes de sustentar o fato de a predição proposta corresponder basicamente a uma versão atrasada da própria série temporal.
- 2.1.3. Embora seja possível obter, com os modelos de predição não-lineares, um desempenho similar àquele do preditor linear (isso não é solicitado ao aluno, no entanto), explique por que o preditor linear é tão competente nesta tarefa de predição específica.
- 2.1.4. Procure justificar também por que alguns preditores não-lineares estão fornecendo uma predição que, aparentemente, falha apenas no *offset* da predição, ou seja, acompanham o comportamento da predição, mas erram no seu valor médio.

Parte 2: Execute o notebook **Q2P2_NLTS.ipynb**, que busca realizar predições de uma série temporal sintética, gerada pelo professor e sem a disponibilidade do processo gerador da série temporal, que garantidamente apresenta comportamento não-linear e estacionário. Atividades práticas:

- 2.2.1. Forneça o notebook devidamente executado, contendo, ao final, um resumo geral do que foi trabalhado em cada seção. Forneça, por exemplo, o número de atrasos da série temporal que está sendo considerado como entrada dos preditores e como está sendo feita a regularização dos modelos, quando for o caso.

- 2.2.2. Dadas as características da série temporal, explique o motivo pelo qual o preditor linear saiu de primeiro colocado em desempenho, na Parte 1, para último colocado nesta Parte 2.
- 2.2.3. Escolha um dos preditores e implemente a predição de múltiplos passos à frente (*several time steps ahead*), explicando como funciona a sua metodologia.
- 2.2.4. Sem a necessidade de implementações práticas, explique que modificações metodológicas você faria caso se constatasse que a série temporal é não-estacionária, mas com uma variação lenta de suas propriedades estatísticas ao longo do tempo.

Parte 3: Tome o preditor linear e ao menos um dos preditores não-lineares implementados e resolva um problema de predição de séries temporais da literatura (predição um passo à frente), à sua escolha, mas envolvendo dados reais e não dados sintéticos. Apresente o notebook com os resultados obtidos, justificando suas decisões metodológicas e analisando os resultados. Se possível, inclua comparações com a literatura. Forneça tudo num notebook **Q2P3.ipynb**.

Apêndice: Conceitos básicos de series temporais

1 Introdução

Em qualquer ramo de atuação profissional, percebe-se um crescimento acentuado na demanda por predições e detecção de tendências junto a variáveis de interesse. Ferramentas computacionais capazes de fornecer predições automáticas acerca de valores futuros de certas variáveis que estão sendo monitoradas já fazem parte do dia-a-dia de muitas empresas privadas e órgãos governamentais, e têm contribuído para o sucesso na definição de políticas estratégicas, em processos de tomada de decisão e em todo tipo de planejamento de curto e médio prazo.

Predição de séries temporais é uma área de pesquisa muito ampla, com desdobramentos na Estatística, na Computação e na Matemática. Serão tratados na Questão 2 apenas alguns aspectos básicos e fundamentais, visando fazer com que o aluno se familiarize com técnicas de síntese de preditores lineares e não-lineares, a partir apenas de dados amostrados das respectivas séries temporais. Outras variáveis de apoio e múltiplas séries correlacionadas poderiam ser consideradas, mas vamos nos restringir ao caso de uma única série temporal por vez. Iremos nos restringir também ao caso de modelos autorregressivos e não iremos recorrer a nenhum outro tipo de tratamento da série temporal, além de aplicar uma operação de escalamento para um intervalo de interesse. Um aprofundamento do estudo pode ser buscado junto à literatura recomendada.

2 Séries temporais e a tarefa de predição

Uma série temporal é dada pelos valores ao longo do tempo de uma variável de interesse, como em:

- ✓ Atividades vitais ou funções orgânicas de um indivíduo;
- ✓ Índices econômicos;
- ✓ Índices sociais;
- ✓ Variáveis climáticas;
- ✓ Variáveis de ecossistemas;
- ✓ Monitoramento de operação de plantas industriais;
- ✓ Monitoramento de processos químicos e físicos.

A frequência de amostragem depende da aplicação e geralmente é fixa. Os intervalos de amostragem mais comuns são ano, mês, semana, dia e hora.

Sendo a predição uma estimativa de valores futuros a partir do conhecimento do histórico de uma variável ou múltiplas variáveis até o presente, surgem algumas questões:

- ✓ O histórico de uma variável até o presente é capaz de auxiliar na predição do seu comportamento futuro?
- ✓ Como geralmente a variável de interesse tem seu comportamento atrelado a uma grande quantidade de fatores e a uma complexa rede de inter-relações, como é possível predizer seu comportamento sem modelar os fenômenos complexos que regem o comportamento da variável e sem monitorar outras variáveis que influenciam nesse comportamento? Exemplo: Como predizer a vazão de um rio com base apenas em seu histórico, sem levar em conta outras variáveis como a vazão dos afluentes, o nível de chuvas na cabeceira do rio, o consumo de água do rio para irrigação agrícola e o grau de assoreamento do leito do rio, dentre outros fatores?

Existem séries temporais que não admitem predição, seja pela independência estatística intrínseca entre o que já se conhece da série e o valor futuro que se pretende predizer, seja pelo fato de que aspectos relevantes do processo estão sendo negligenciados, como em séries multidimensionais em que alguma(s) variável(is) não está(ão) sendo monitorada(s). Nesses casos, geralmente as iniciativas de predição tendem a produzir resultados decepcionantes.

Por outro lado, existem séries temporais que apresentam uma dependência estatística significativa entre o histórico da série e o valor futuro a ser predito. Nesses casos, as predições tendem a produzir resultados de predição de alta qualidade, mesmo que não se conheçam especificidades do processo que gera a série temporal.

Esta dependência estatística (entre o histórico da série e o valor a ser predito) pode ser estimada a priori. Com isso, antes mesmo de se iniciar o projeto do preditor já é possível ter uma boa noção de como pode vir a ser o seu desempenho naquela tarefa de predição. Além disso, análises de dependência estatística auxiliam na definição da estrutura do preditor, mais especificamente na definição do vetor de regressão, ou seja,

o vetor de entrada do preditor, também chamado de linha de derivação de atrasos. Para tanto, geralmente são respondidas duas perguntas:

- Qual é o tamanho da janela de valores passados a serem considerados na entrada do preditor?
- Quais valores passados dentro desta janela devem ser considerados? Todos eles ou só alguns deles?

Coefficientes de correlação linear (contextualizados por Francis Galton em 1885 e devidamente formalizados por Karl Pearson em 1895) (RODGERS & NICEWANDER, 1988; STIGLER, 1989) são muito utilizados para responder a essas perguntas, mas cabe salientar que eles supõem o uso subsequente de preditores lineares. Um índice mais consistente, capaz de capturar dependências lineares e não-lineares, é a informação mútua, derivado da Teoria de Informação (COVER, 1991; REZA, 1994; WESTIAN, 1990).

3 Metodologia

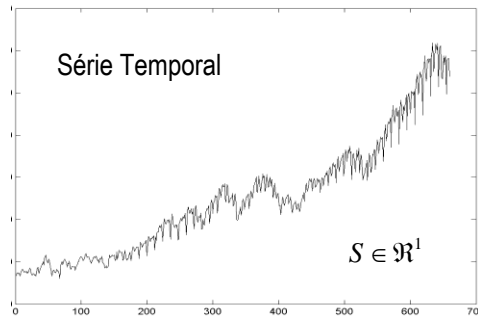
As primeiras tentativas no campo da predição de séries temporais foram efetuadas nos anos 20, quando YULE (1927) aplicou um modelo autorregressivo linear no estudo de manchas solares. Nos anos 50, DOOB (1953) prosseguiu a investigação com a análise teórica de séries temporais estacionárias. Já nos anos 70, foram propostas as técnicas e metodologias que obtiveram maior destaque a partir de então, reunidas no trabalho de BOX & JENKINS (1976).

Os métodos de Box & Jenkins baseiam-se na proposição de que o valor atual da série temporal é a combinação de p valores precedentes e q impactos aleatórios anteriores, mais o impacto atual. Os p valores precedentes formam o *componente autorregressivo* e os q impactos prévios formam o *componente de média móvel* da série. Obtêm-se assim os bem conhecidos modelos ARMA (do inglês *autoregressive moving average*). A modelagem de uma série temporal tem por objetivo, então, a determinação dos valores de p e q , seguida da estimação dos respectivos coeficientes da combinação linear. Gera-se, assim, um problema de regressão linear que tem como uma de suas formulações aquela que leva à solução de quadrados mínimos.

Nos últimos anos, considerável atenção tem sido dedicada a métodos alternativos para o estudo de séries com dependências não-lineares, destacando-se a utilização de redes neurais artificiais. O emprego das arquiteturas MLP e RBF, ao longo das últimas décadas, trouxe resultados muito positivos no campo da predição de valores futuros em séries temporais, em virtude do caráter essencialmente não-linear dessas estruturas.

Para o emprego de uma rede neural artificial como preditor de um passo à frente, é necessário definir quais valores passados da série serão utilizados na definição da entrada da rede neural. Feito isso, o problema de síntese do preditor se transforma em um problema de treinamento supervisionado, onde o que se deseja é obter um mapeamento multidimensional não-linear de entrada-saída, como indicado na sequência de passos abaixo.

Passo 1: Obter a série temporal, ou seja, os valores históricos da variável a ser predita um passo à frente. Se necessário, faça um escalamento dos dados, evitando que o intervalo de excursão dos valores seja qualquer. Outros tipos de pré-processamento, como diferenciar a série temporal, também podem ser considerados visando eliminar tendências e sazonalidades, por exemplo.



$$S = s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8 \dots, s_N$$

Passo 2: Definir quais valores passados da série serão considerados na predição. Suponha aqui que L valores passados consecutivos sejam considerados. Com isso, monte a tabela a seguir, a qual retrata o comportamento desejado do preditor.

s_{t-L+1}	s_{t-L+2}	...	s_t	s_{t+1}
s_1	s_2	...	s_L	s_{L+1}
s_2	s_3	...	s_{L+1}	s_{L+2}
\vdots	\vdots	\vdots	\vdots	\vdots
s_{N-L}	s_{N-L-1}	...	s_{N-1}	s_N
\Downarrow	\Downarrow		\Downarrow	\Downarrow
X_1	X_2		X_L	Y

Passo 3: Separe os dados da tabela acima em 3 conjuntos: conjunto de treinamento, conjunto de validação e conjunto de teste. Não é necessário que a separação seja sequencial, mas a separação sequencial é mandatória no caso de redes neurais recorrentes, razão pela qual será adotada aqui.

Passo 4: Treine a rede neural com o conjunto de treinamento (ela vai produzir um mapeamento do \mathcal{R}^L para o \mathcal{R}^1) e pare o treinamento quando for atingido o valor mínimo do erro quadrático médio para o conjunto de validação.

Passo 5: Avalie o preditor recém-obtido junto aos dados de teste.

4 Agradecimento

Algumas figuras e trechos de texto utilizados neste apêndice são de autoria de Wilfredo Jaime Puma Villanueva, que realizou a sua pós-graduação junto ao Programa de Pós-Graduação da FEEC/Unicamp.

5 Referências

BOX, G. E. P. & JENKINS, G. M. (1976). *Time Series Analysis, Forecasting and Control*. Holden Day.

BRADLEY, E. & KANTZ, H. (2015). *Nonlinear time-series analysis revisited*. arXiv:1503.07493v1.

COVER, T. M. & THOMAS, J. A. (1991) *Elements of Information Theory*, Wiley.

DOOB, J. (1953). *Stochastic Processes*. Wiley.

PRECHELT L. (1997). Early Stopping - but when?, Technical Report URL: http://wwwipd.ira.uka.de/~prechelt/Biblio/stop_tricks1997.ps.gz

PRECHELT L. (1998). Automatic early stopping using cross validation: Quantifying the criteria. *Neural Networks*, vol. 11, no 4, pp. 761-767.

REZA, F. M. (1994) *An Introduction to Information Theory*, Dover.

RODGERS, J. L. & NICEWANDER, W. A. (1988) Thirteen ways to look at the correlation coefficient. *The American Statistician*, vol. 42, no. 1, pp. 59-66.

STIGLER, S. M. (1989). Francis Galton's Account of the Invention of Correlation. *Statistical Science*, vol. 4, no. 2, pp 73-79.

WEIGEND, A. S. & GERSHENFELD, N. A. (1993). *Time Series Prediction: Forecasting the Future and Understanding the Past*. Perseus Press.

WENTIAN, L. (1990) Mutual information functions versus correlation functions. *Journal of Statistical Physics*, vol. 60, nos. 5-6, pp. 823-837.

YULE G. (1927). On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers. *Philos. Trans. R. Soci.*, A226.