



Universidade do Minho

Departamento de Informática

Mestrado [integrado] em Engenharia Informática

Perfil de Machine Learning: Fundamentos e Aplicações

Classificadores e Sistemas Conexionistas

1º/4º Ano, 2º Semestre

Ano letivo 2020/2021

Trabalho Prático de Grupo – 2ª Parte

Abril, 2021

Tema	Benchmarking de modelos de <i>Deep Learning</i> para séries temporais.
Objetivos de aprendizagem	Com a realização deste trabalho prático pretende-se sensibilizar e motivar os alunos para a conceção e desenvolvimento de modelos de <i>Deep Learning</i> nos mais variados domínios, com particular ênfase no que respeita à construção de modelos capazes de trabalhar séries temporais.
Enunciado	<p>Durante muitos anos, as previsões em séries temporais eram obtidas recorrendo, essencialmente, a modelos estatísticos. Contudo, nos últimos anos, têm surgido vários estudos e trabalhos que demonstram a capacidade de modelos de <i>Deep Learning</i> em produzir melhores previsões neste domínio. Na sua essência, uma série temporal é uma sequência ordenada de dados. Exemplos incluem, entre muitos outros, dados do mercado de ações, fluxo de tráfego e até mesmo texto.</p> <p>São várias as possibilidades quando se discute a conceção de modelos de <i>Deep Learning</i> para séries temporais. Começando em redes mais simples, como as <i>Multi-Layer Perceptrons</i>, até às baseados em recorrência, convolucionais e até mesmo híbridas, as hipóteses são vastas.</p> <p>No contexto da componente prática de avaliação desta unidade curricular, pretende-se que os grupos de trabalho façam o benchmarking de modelos de <i>Deep Learning</i> para previsão de séries temporais. Para isso, deverão trabalhar o <i>dataset</i> disponível em http://bit.ly/3kmGpZN. Para além do referido <i>dataset</i>, deverão também procurar e selecionar, pelo menos, mais um domínio (i.e., <i>dataset</i>) sobre os quais farão as avaliações dos modelos, i.e., deverão conceber e otimizar os modelos para cada <i>dataset</i>, mantendo as condições imutáveis entre modelos no mesmo <i>dataset</i>. Os grupos deverão seguir uma, de várias, metodologias existentes para planeamento e execução de projetos de <i>Machine Learning</i>. Neste trabalho prático, e como em qualquer projeto de <i>Machine Learning</i>, a exploração, tratamento e processamento dos dados assume especial relevo.</p> <p>Os modelos a desenvolver deverão ser definidos em função da configuração do problema a tratar, i.e., modelos <i>single-step</i> vs <i>multi-step</i>, modelos univariável vs multivariável e a forma (<i>shape</i>) do input. A otimização do modelo deve focar a sua estrutura e os hiper-parâmetros definidos, utilizando, para isso, técnicas como intuição e experimentação, <i>random search</i> ou algoritmia genética, por exemplo. A performance do modelo deve ser validada utilizando <i>cross-validation</i> para séries temporais.</p> <p>Os resultados obtidos deverão ser objeto de um relatório que contenha, entre outros:</p> <ul style="list-style-type: none">• Quais os domínios a tratar, quais os objetivos e como se propõe a atingi-los;• Qual a metodologia seguida e como foi aplicada;• Análise detalhada dos <i>datasets</i> trabalhados e de todo e qualquer tratamento efetuado;• Descrição da configuração e enquadramento adotado para o problema em causa;• Descrição detalhada dos modelos desenvolvidos, quais as suas características, como e sobre que parâmetros foi realizado o <i>tuning</i> do modelo, características do treino, entre outros detalhes que seja oportuno fornecer;

- Sumário dos resultados obtidos e respetiva análise crítica.

Todo o processo deverá ser acompanhado de exemplos e indicações que permitam reproduzir todos os passos realizados assim como os resultados obtidos.

Entrega e avaliação

As submissões deverão ser feitas na plataforma de e-learning da Universidade do Minho, em "*Conteúdo/Instrumentos de Avaliação em Grupo/Submissão TP2*", enviando, num único ficheiro compactado, todos os elementos produzidos. As submissões deverão ser realizadas até ao final do dia 16 de maio de 2021.

A sessão de apresentação do trabalho decorrerá no dia 20 de maio de 2021, tendo início às 13h00min. Cada grupo disporá de 10 minutos para realizar a apresentação, utilizando os meios que considerar mais adequados.

Referências bibliográficas

Para além do material disponibilizado nas aulas aconselha-se também a consulta das seguintes fontes:

- Hochreiter, S. & Schmidhuber, J., Long Short-Term Memory. *Neural Computation*. 9.8 (1997), pp. 1735–1780. doi: 10.1162/neco.1997.9.8.1735.
- Fernandes, B., Silva, F., Alaiz-Moretón, H., Novais, P., Neves, J. & Analide, C., Long Short-Term Memory Networks for Traffic Flow Forecasting: Exploring Input Variables, Time Frames and Multi-Step Approaches. *INFORMATICA*, Vol. 31(4), pp. 723-749, 2020. doi: 10.15388/20-INFOR431.
- Multi-step Time Series Forecasting with Long Short-Term Memory Networks in Python, Jason Brownlee, 2017.