

Professor: Rafael Stubs Parpinelli

Estagiário Docente: Douglas Macedo Sgrott

Data de lançamento: 16/06/2021 - 18:00

Data prevista de término: 21/06/2021 - 23:59

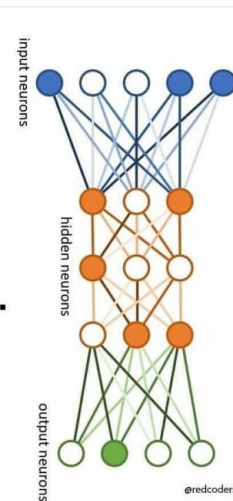
Disciplina: Inteligencia Artificial

Trabalho Final - Modelagem de uma Rede Neural (python)

**THIS IS A NEURAL
NETWORK.**

**IT MAKES MISTAKES.
IT LEARNS FROM THEM.**

**BE LIKE A NEURAL
NETWORK.**



Objetivos:

- Depois de limpar e pré-processar um dataset, criar uma modelagem de dados usando Redes Neurais.
- Entender como diferentes pré-processamentos e parâmetros da rede neural podem influenciar a performance da modelagem.
- Unir os conhecimentos dos trabalhos de EDA (trabalho 1), parametrização de redes neurais (trabalho 2) e métricas de regressão (trabalho 3) em um trabalho final.

Sobre o dataset:

- **Variáveis de entrada / independentes:** 'city', 'area', 'rooms', 'bathroom', 'parking spaces', 'floor', 'animal', 'furniture', 'hoa (R\$)', 'rent amount (R\$)', 'property tax (R\$)', 'fire insurance (R\$)'
- **Variável de saída / dependente:** 'total (R\$)'

Como entregar o trabalho:

- Faça um fork/cópia do notebook deste endereço:
<https://www.kaggle.com/dougsgrott/ia-trabalho-final?scriptVersionId=65962271>
- Implemente seus códigos. Quando quiser salvar os resultados, clique em Save Version -> Quick Save
- Quando finalizar, disponibilize o notebook com visibilidade pública (ou convide o estagiário para ser colaborador)
- Envie o link do notebook para o estagiário (email: doug.sgrott@gmail.com ou whatsapp, número presente no grupo da disciplina). O trabalho será apenas avaliado no fim do prazo estipulado.

O que será avaliado de forma objetiva:

- Se o modelo está ajustado (sem overfit/underfit). Ou seja, se as curvas de treino e validação estão suficientemente próximos.
- A quantidade de dados (linhas de df) e features (colunas de df) foram utilizados.
- O valor de R^2

O que será avaliado de forma subjetiva:

- Esforço. Ou seja, quando mais explorar as combinações de pre-processamento + parâmetros de rede, melhor. Para isso, crie novas versões do notebook.

Observações:

- O valor de R será analisado juntamente com o número de features e dados usados. Ou seja, um R alto obtido usando poucos dados e features não vale muita coisa. Mas é melhor que nada ;)
- Se deixou de utilizar colunas do df = menos pontos. Se fez engenharia de features = mais pontos (pouco). Isto serve para incentivar a utilizar os dados categóricos.

Dicas:

- Quando alcançar um resultado que queira salvar, clique em Save Version -> Quick Save. Isto fará com que o seus resultados atuais sejam mostrados na versão html. Se clicar em Save Version -> Save & Run All (commit), o notebook será rodado novamente, e isso pode modificar os resultados obtidos.
- No começo, KISS (Keep it simple, silly). Comece de forma simples. Comece com menos features e com pouco pre processamento. Recomendo começar usando variáveis numéricas e filtro de outliers. Isso já deve ser suficiente para obter um R alto. MAS ISSO NÃO É SUFICIENTE! Depois que conseguir um modelo com alto R usando um modelo e pre processamento simples, salve a versão do notebook e continue explorando! A meta é conseguir utilizar TODAS as features e o maior número de dados possível. O objetivo é explorar a combinação de pré-processamento de dados + parâmetros da rede neural.
- Eu vou analisar as diferentes versões do notebook (se houver). Ou seja, se na primeira versão obteve um modelo bom usando poucas features, ok... Se na segunda versão obteve um modelo um pouco pior ou melhor usando mais features, melhor ainda. Explore as combinações! You're a wizard Harry!

O notebook está organizado em partes:

- **Dataset:** Onde você irá limpar e pre processar o dataset. Atribua a versão final do dataset em um dataframe chamado **df**.
- Separação dos dados: Aqui os dados são normalizados e divididos em Treino/Validação. Não precisa modificar o código.
- **Arquitetura da Rede Neural:** Onde você vai definir a arquitetura da rede neural.
- **Parâmetros de otimização da Rede Neural:** Onde você vai definir outros parâmetros da rede neural.
- Visualização dos resultados: Onde os resultados são obtidos Exemplos: Servir como exemplo de análise, data cleaning e pré-processamento.
 - O que está em **negrito**, DEVE ser modificado pelo aluno. Modificações em 'Separação dos dados' e 'Visualização dos resultados' podem penalizar o trabalho.