

Trabalho Final: Redes neurais e geociências

O objetivo deste trabalho é utilizar o framework Pytorch para desenvolver diferentes arquiteturas de redes neurais e compará-las em problema de classificação de granulometria.

Para o desenvolvimento do trabalho, torna-se necessário cumprir os seguintes pré-requisitos:

- Ser feito individualmente;
- [Baixar os dados](#), criar código Jupyter utilizando o Pytorch e documentá-lo seguindo o processo de desenvolvimento, disponibilizar o código e o arquivo resultante do dataset hidden.

Data de entrega e apresentação: **02/07/2021**.

1. Informações do trabalho:

A base de dados que será utilizada nesse trabalho contém informações geológicas com etapas de pré-processamento já executadas. Foi criada por um time de especialistas em litologia e contém informações de sensores e 11 classes a granulométricas na coluna "FORCE_2020_LITHOFACIES_LITHOLOGY":

- 30000: Sandstone
- 65030: Sandstone/Shale
- 65000: Shale
- 80000: Marl
- 74000: Dolomite
- 70000: Limestone
- 70032: Chalk
- 88000: Halite
- 86000: Anhydrite
- 99000: Tuff
- 90000: Coal
- 93000: Basement

Para este trabalho, **você deve desenvolver:**

- No mínimo 3 diferentes arquiteturas de redes neurais, contendo diferentes configurações de parâmetros e hiperparâmetros;
- Considere utilizar diferentes quantidades de neurônios, camadas, métodos de inicialização, métodos de otimização e cálculo de gradiente;
- Avalie overfitting e underfitting, utilizando Split de treinamento, teste e validação;
- Documente toda a etapa de desenvolvimento;
- Considere acurácia para validar os resultados.

2. Entrega:

- I. **Código(s) desenvolvido para resolução do problema devidamente documentado utilizando a base lithology.csv;**
- II. **Resultado da classificação do arquivo hidden.csv:**

- Será divulgado um placar dos resultados do trabalho entre todos os alunos, porém seu resultado não será utilizado como critério de avaliação, apenas para que você tenha uma ideia do resultado comparado ao de seus colegas.
- O arquivo a ser submetido em formato “CSV” conterá uma única coluna “lithology” com exatamente o mesmo número de linhas que dataset hidden.

Exemplo:

lithology

65000

65000

30000

30000

30000

30000

30000

65030

65030

65030

.

.

.

3. Critérios de Avaliação:

- Resolução adequada, cumprindo os pré-requisitos do problema e utilizando as ferramentas e técnicas apropriadas;
- Redação correta, organizada e fluente do relatório.
- Avaliação e documentação dos resultados obtidos.
- Cumprimento do prazo de entrega. Cada semana de atraso na entrega acarreta 2 pontos a menos na nota do trabalho.

Observações: Podem ser executadas etapas de análise de dados, utilizar menos features e menos dados para treinamento, desde que isso seja justificado no trabalho. Soluções interessantes do problema receberão pontos a mais durante a avaliação.