Tópicos em Aprendizado de Máquina e Otimização 2024-2 Trabalho 1

Utilizando a base de dados *Diagnostic Wisconsin Breast Cancer Database.*, mostrada na figura abaixo, disponível em

https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic.pede-se:

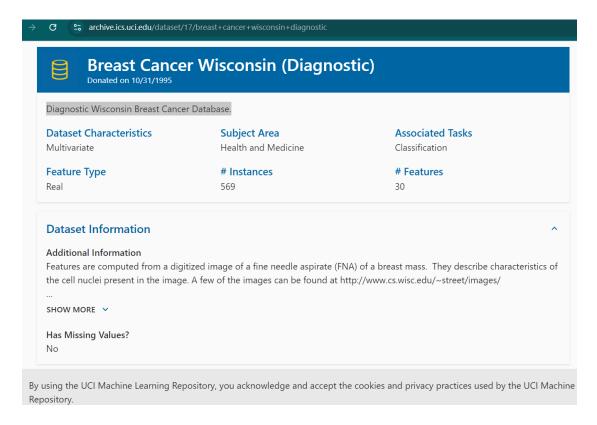
- 1) Utilize 80% dos dados para o treinamento, 20% para validação e 20% para teste. Obtenha esses conjuntos de forma aleatória;
- 2) Treine uma rede neural com duas camadas de neurônios nas seguintes condições:
  - a. sem regularização;
  - b. sem regularização e com para da antecipada com paciência igual a 3;
  - c. com regularização L2 nas duas camadas;
  - d. com regularização L<sub>1</sub> nas duas camadas;
  - e. com dropout de 30% na segunda camada de neurônios
  - f.  $com\ dropout\ de\ 30\%$  na segunda camada de neurônios e regularização  $L_2$  em todas as camadas.
- 3) Em todas as simulações fixe um número máximo de épocas para o treinamento. Na letra b, devido a parada antecipada, o treinamento pode treinar antes desse número máximo. Nas outras simulações, o treinamento terminará sempre nesse número máximo. Para fixar esse número máximo, faça uma simulação inicial sem regularização e sem parada antecipada e veja em que época o erro no conjunto de validação estabiliza.
- 4) Em todos os treinamentos, à exceção do modelo com parada antecipada, avalie o resultado no conjunto de teste utilizando o modelo com melhor valor de acurácia no conjunto de validação. Com o tensorflow+keras isso pode ser feito com a classe "callback"

```
EPOCHS = 10
checkpoint_filepath = '/tmp/ckpt/checkpoint.model.keras'
model_checkpoint_callback = keras.callbacks.ModelCheckpoint(
    filepath=checkpoint_filepath,
    monitor='val_accuracy',
    mode='max',
    save_best_only=True)
```

# Model is saved at the end of every epoch, if it's the best seen so far. model.fit(epochs=EPOCHS, callbacks=[model\_checkpoint\_callback])

# The model (that are considered the best) can be loaded as - keras.models.load model(checkpoint filepath)

- 5) Use o conjunto de validação para escolher a melhor dentre as opções treinadas;
- 6) Avalie o desempenho da melhor opção treinada no conjunto de teste.
- 7) Caso esteja havendo *underfitting*, pode aumentar mais uma camada de neurônios, para avaliar o comportamento.



Obervações: 1) Acrescente um link para o código; 2) o relatório deve seguir o modelo dos relatórios do curso 1; 3) O número máximo de páginas é 3. Não ultrapasse!!!

Data: 11/09/2024