

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 parada antecipada com paciência igual a 3;
  - c. com regularização  $L_2$  nas duas camadas;
  - d. com regularização  $L_1$  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.

   archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic



## Breast Cancer Wisconsin (Diagnostic)

Donated on 10/31/1995

Diagnostic Wisconsin Breast Cancer Database.

Dataset Characteristics

Multivariate

Feature Type

Real

Subject Area

Health and Medicine

# Instances

569

Associated Tasks

Classification

# Features

30

Dataset Information

Additional Information

Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image. A few of the images can be found at <http://www.cs.wisc.edu/~street/images/>

...

SHOW MORE

Has Missing Values?

No

By using the UCI Machine Learning Repository, you acknowledge and accept the cookies and privacy practices used by the UCI Machine Repository.

Observações: 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