

TP555 - AI/ML

Lista de Exercícios #12

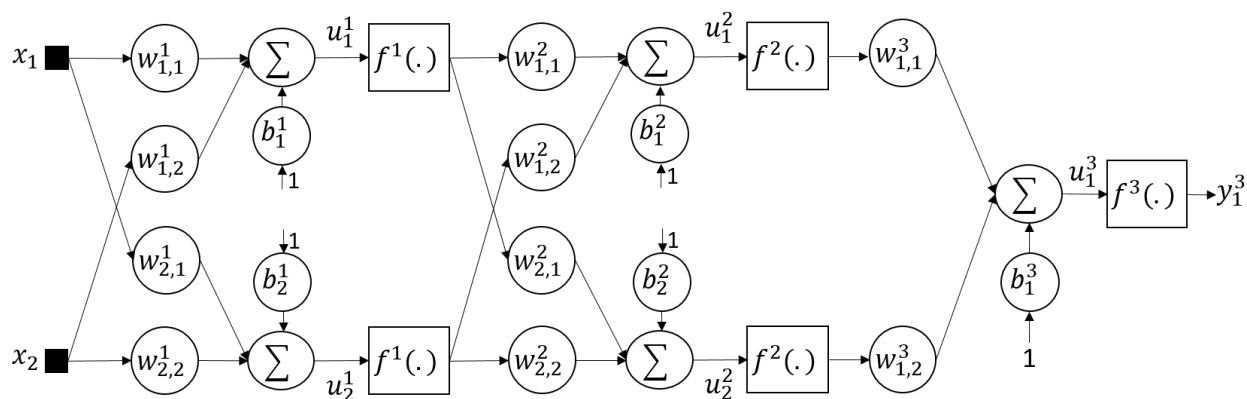
Redes Neurais Artificiais (Parte 2)

1. Por que a função de ativação logística foi um ingrediente chave no treinamento das primeiras redes MLPs?
2. Utilizando o exemplo `activation_functions.ipynb` como base, cite três funções de ativação diferentes das que vimos. Plote a função e sua respectiva derivada.
3. Desenhe uma rede MLP que calcule $A \oplus B$, onde \oplus representa a operação lógica XOR. (Dica: $A \oplus B = (A \& !B) \mid (!A \& B)$)
4. Suponha que você tenha uma rede MLP composta por uma camada de entrada com 10 neurônios, seguida por uma camada oculta com 50 neurônios e, finalmente, uma camada de saída com 3 neurônios. Todos os neurônios usam a função de ativação ReLU. Sendo assim, responda
 - a. Qual é a dimensão da matriz de entrada X ?
 - b. Qual a dimensão do vetor de pesos, W_h , da camada oculta e de seu vetor de bias b_h ?
 - c. Qual é a dimensão do vetor de pesos, W_o , da camada de saída e de seu vetor de bias b_o ?
 - d. Qual é a dimensão da matriz de saída, Y , da rede?
 - e. Escreva a equação que calcula a matriz de saída da rede, Y , como uma função de X , W_h , b_h , W_o e b_o .
5. Quantos neurônios você precisa na camada de saída de uma rede MLP se você deseja classificar emails em spam ou ham? Qual função de ativação você deve usar na camada de saída? Se você deseja classificar a base de dados MNIST (imagens de dígitos escritos à mão), quantos neurônios você precisa na camada de saída usando qual tipo de função de ativação?
6. Liste todos os hiperparâmetros que você pode ajustar em uma rede MLP? Caso você perceba que a rede MLP está **sobreajustando**, como você pode modificar esses hiperparâmetros para tentar resolver o problema?

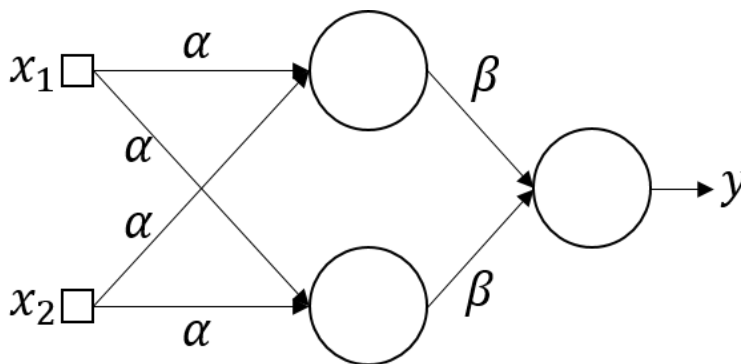
7. Dada a rede MLP mostrada na figura abaixo, encontre a derivada parcial do erro de saída, $e = d - y_1^3$, em relação aos dois pesos abaixo. (**Dica:** Siga o equacionamento apresentado no material da aula.)

a. $w_{1,1}^1$, ou seja, $\frac{\partial e}{\partial w_{1,1}^1}$

b. $w_{2,2}^1$, ou seja, $\frac{\partial e}{\partial w_{2,2}^1}$

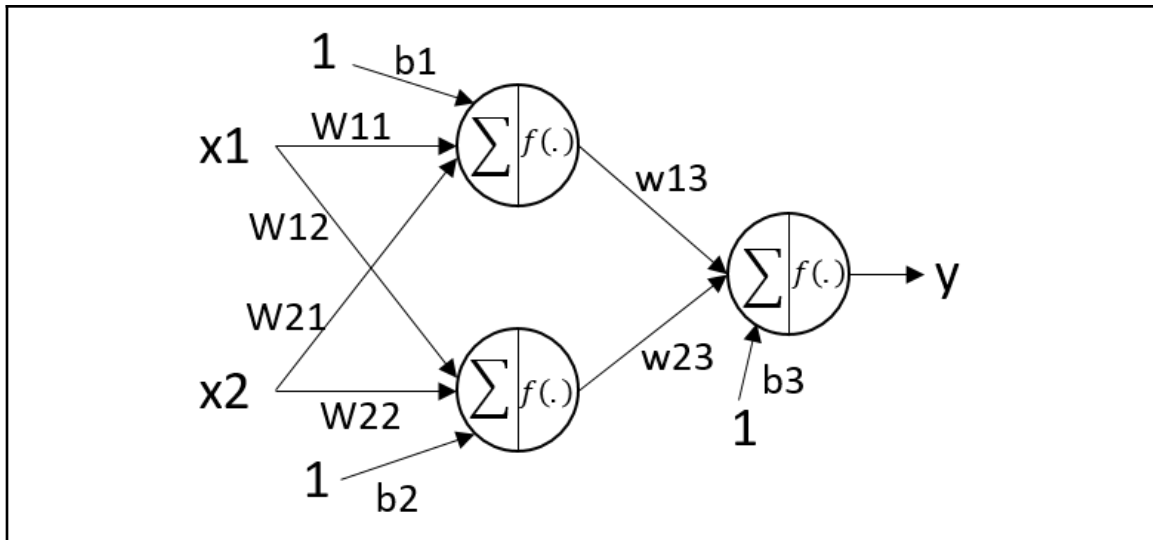


8. O que aconteceria com os valores dos gradientes e pesos quando se inicializa **todos** os pesos de uma rede neural com o mesmo valor constante? Para verificar o que aconteceria, considere uma rede neural com dois atributos de entrada, (x_1, x_2) , e com dois nós ocultos (ou seja, uma camada escondida com 2 nós) com função de ativação sigmóide e assuma que inicializamos todos os pesos de bias com o valor 0 e os pesos com alguma constante α . Esta rede possui um único nó na camada de saída com função de ativação sigmóide e pesos de bias inicializados com o valor 0 e os pesos inicializados com alguma constante β . A rede neural é apresentada na figura abaixo. Agora, aplique uma entrada (x_1, x_2) e calcule o erro de saída, $e = d - y$, onde d é o valor esperado. Na sequência, use o algoritmo da retropropagação do erro para calcular os gradientes de todos os pesos da rede neural.



- a. O que pode ser dito com relação à influência da saída dos nós ocultos no erro de saída?

- b. O que ocorre com os gradientes dos nós da camada oculta (Dica: compare os gradientes de ambos os nós da camada oculta.)?
 - c. O que pode ser feito para resolver este comportamento que você observou quando os pesos são inicializados com o mesmo valor constante?
9. Baseando-se no exemplo `MLPWithTensorFlowLowLevelAPI.ipynb`, utilize objetos da classe ***Dense()*** do módulo ***tf.keras.layers*** ao invés da função ***neuron_layer()***. Após treinar o modelo, você percebeu alguma diferença na performance do seu modelo? Se sim, qual foi esta diferença?
(Dica: A documentação da classe ***Dense*** pode ser encontrada no seguinte link: https://www.tensorflow.org/api_docs/python/tf/keras/layers/Dense)
(Dica: Lembre que ***Dense*** é uma classe que precisa ser instanciada. Apenas após a criação do objeto da classe ***Dense*** é que se deve passar a entrada para esta camada de nós.)
10. Baseando-se no código do exercício anterior, implemente a técnica do ***early stop*** de forma que o modelo sempre armazene os pesos que resultem no menor erro. Além disso, crie um contador que conte o número de épocas sem progresso, ou seja, o número de vezes em que o erro da época não diminuiu com relação ao menor erro obtido até o momento. Com base nesse contador, faça com que o treinamento se encerre caso o contador atinja 50 épocas sem progresso.
(Dica: Lembre-se que uma maneira de se evitar que um modelo ***sobreajuste*** é interromper o treinamento assim que o erro de validação/teste atinja o mínimo. Outra maneira válida é armazenar os pesos do modelo que resultem no menor erro de validação durante o treinamento através de um pré-determinado número de épocas. Essa abordagem é chamada de ***early stop***.)
11. Baseando-se no código do exercício anterior, utilize o otimizador que implementa o algoritmo momentum ao invés do gradiente descendente e treine uma rede MLP que obtenha uma precisão maior do que 98%. Use o trecho de código abaixo.
- ```
optimizer = tf.train.MomentumOptimizer(learning_rate=0.01, momentum=0.9)
```
12. **Exercício sobre multi-layer perceptrons com TensorFlow:** Utilizando o TensorFlow, implemente um Multi Layer Perceptron (MLP) com 3 unidades (i.e., perceptrons), mostrada na figura abaixo, que classifique a função lógica XOR.



Se nós imaginarmos este classificador MLP da figura em forma matricial, então nós temos a seguinte equação:

$$y = f(f(W \cdot X + B) \cdot w + b3),$$

onde:

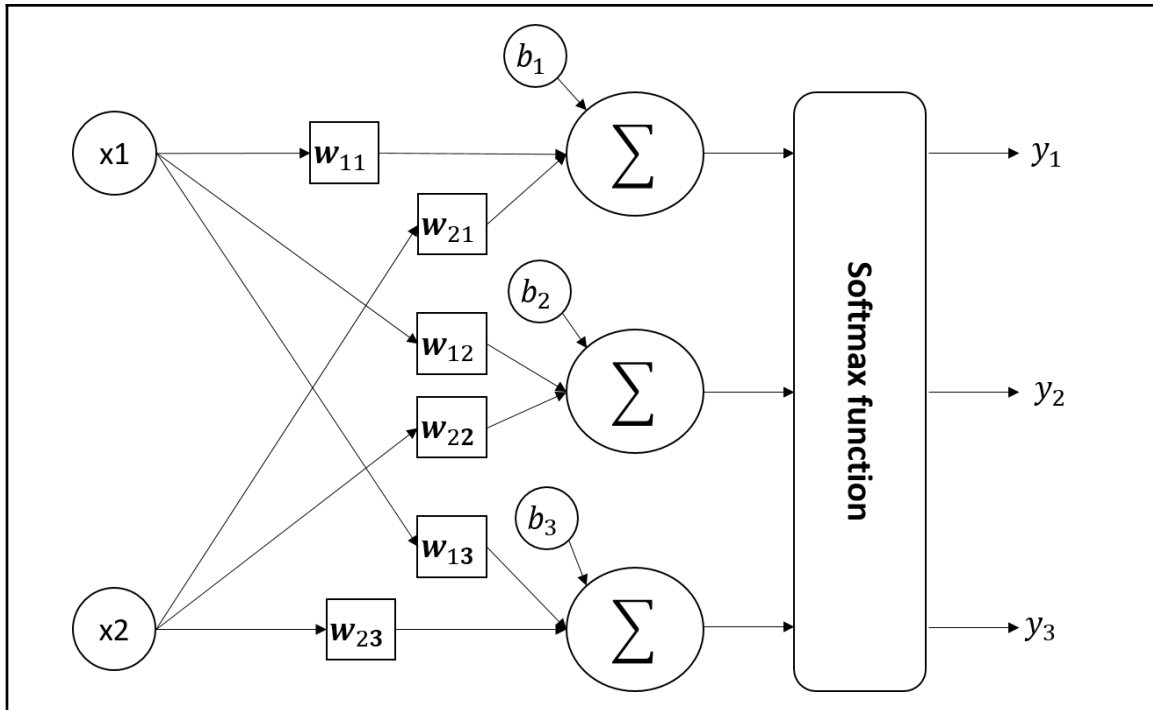
- X é o vetor de entrada com dimensão 2x1.
- W é uma matriz de coeficientes para a primeira camada da MLP, com dimensão 2x2.
- B é um vetor de bias para a primeira camada, com dimensão 2x1.
- w é um vetor de coeficientes para a segunda camada da MLP, com dimensão 2x1.
- b3 é um escalar de bias para a segunda camada, com dimensão, 1x1.
- y é o valor de saída da MLP.
- f(.) é a função de ativação sigmóide.

Use 40000 épocas e imprima o erro a cada 100 épocas, em seguida, faça o seguinte:

- a) Crie um grafo utilizando o **GradientDescentOptimizer** para classificar a função lógica XOR.  
(**Dica:** verifique a documentação deste otimizador no seguinte link: [https://www.tensorflow.org/api\\_docs/python/tf/compat/v1/train/GradientDescentOptimizer](https://www.tensorflow.org/api_docs/python/tf/compat/v1/train/GradientDescentOptimizer))
- b) Imprima e salve o erro (i.e., loss) a cada 100 épocas, por exemplo.
- c) Treine o modelo.
- d) Plote um gráfico mostrando o erro versus o número de épocas.
- e) Plote um gráfico com as fronteiras de decisão.
- f) Crie um segundo grafo utilizando o **AdamOptimizer** para classificar a função lógica XOR.  
(**Dica:** Não se esqueça de resetar o grafo com **tf.reset\_default\_graph().**)  
(**Dica:** verifique a documentação deste otimizador no seguinte link: [https://www.tensorflow.org/api\\_docs/python/tf/compat/v1/train/AdamOptimizer](https://www.tensorflow.org/api_docs/python/tf/compat/v1/train/AdamOptimizer))
- g) Imprima e salve o erro (i.e., loss) a cada 100 épocas, por exemplo.

- h) Treine o modelo.
- i) Plote um gráfico mostrando o erro versus o número de épocas.
- j) Plote um gráfico com as fronteiras de decisão.
- k) Baseado nos valores de precisão calculados anteriormente, qual dos 2 otimizadores tem melhor performance?

13. Utilizando o TensorFlow, implemente o modelo mostrado na figura abaixo, para classificar 3 classes diferentes.



Se nós modelarmos o classificador da figura em formato matricial, então, nós temos a seguinte equação:

$$\mathbf{y} = f(\mathbf{X} \cdot \mathbf{W} + \mathbf{b}),$$

onde:

- $\mathbf{X}$  é a matriz de entrada com dimensão  $N \times 2$ .
- $\mathbf{W}$  é a matriz de coeficientes com dimensão  $2 \times 3$ .
- $\mathbf{b}$  é o vetor de bias para a primeira camada, com dimensão 3.
- $\mathbf{y}$  é o vetor de saída do classificador.
- $f(\cdot)$  é a função softmax.
- $N$  é o número de exemplos.

Crie a base de dados das 3 classes com o código abaixo.

```
from sklearn.datasets import make_blobs

Create a 3-class dataset for classification
N = 1000
centers = [[-5, 0], [0, 1.5], [5, -1]]
X_, y_ = make_blobs(n_samples=N, centers=centers, random_state=42)
```

Em seguida, faça o seguinte:

- a) Plote um gráfico mostrando as diferentes classes. (**Dica:** use cores ou marcadores diferentes para cada classe.)
- b) Crie um grafo utilizando o **GradientDescentOptimizer** para classificar os dados. (**Dica:** verifique a documentação deste otimizador no seguinte link: [https://www.tensorflow.org/api\\_docs/python/tf/compat/v1/train/GradientDescentOptimizer](https://www.tensorflow.org/api_docs/python/tf/compat/v1/train/GradientDescentOptimizer))
- c) Imprima e salve o erro e a precisão a cada 100 épocas.
- d) Treine o modelo. (**Dica:** Use 20000 épocas ou treine até que a precisão seja de 100%.)
- e) Plote um gráfico mostrando o erro e a precisão versus o número de épocas.
- f) Plote um gráfico com as fronteiras de decisão.
- g) Crie um segundo grafo utilizando o **AdamOptimizer** para classificar os dados. (**Dica:** Não se esqueça de resetar o grafo com **tf.reset\_default\_graph()**.) (**Dica:** verifique a documentação deste otimizador no seguinte link: [https://www.tensorflow.org/api\\_docs/python/tf/compat/v1/train/AdamOptimizer](https://www.tensorflow.org/api_docs/python/tf/compat/v1/train/AdamOptimizer))
- h) Imprima e salve o erro e a precisão a cada 100 épocas..
- i) Treine o modelo. (**Dica:** Use 20000 épocas ou treine até que a precisão seja de 100%.)
- j) Plote um gráfico mostrando o erro e a precisão versus o número de épocas.
- k) Plote um gráfico com as fronteiras de decisão.
- l) Baseado nos valores de erro e precisão impressos e plotados nas figuras anteriores, qual dos 2 otimizadores tem melhor performance? (**Dica:** Qual dos 2 converge mais rapidamente?)

**OBSERVAÇÃO:** Neste exercício, você vai precisar utilizar a função **tf.nn.sparse\_softmax\_cross\_entropy\_with\_logits()** para calcular o erro do modelo e assim, conseguir treiná-lo, ou seja, encontrar os pesos que minimizem o erro. A função **tf.nn.sparse\_softmax\_cross\_entropy\_with\_logits()** é equivalente a aplicar a função de ativação softmax e, em seguida, calcular a entropia cruzada, mas é mais eficiente e cuida adequadamente de casos incomuns, por exemplo, onde os logits são iguais a 0. A documentação dessa função pode ser encontrada em:

[https://www.tensorflow.org/api\\_docs/python/tf/nn/sparse\\_softmax\\_cross\\_entropy\\_with\\_logits](https://www.tensorflow.org/api_docs/python/tf/nn/sparse_softmax_cross_entropy_with_logits)

Seguem alguns exemplos de como utilizar a função:

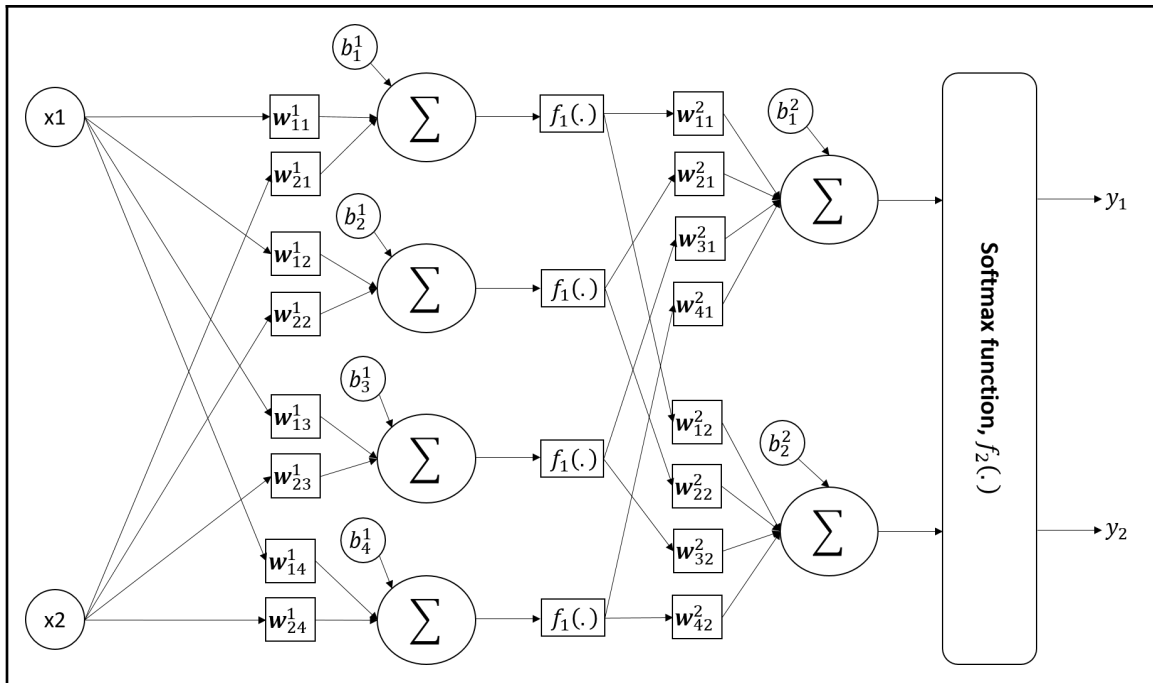
[1]

<https://towardsdatascience.com/active-learning-on-mnist-saving-on-labeling-f3971994c7ba>

[2]

<https://riptutorial.com/tensorflow/example/17628/computing-costs-on-a-softmax-output-layer>

14. **Exercício sobre TensorFlow:** Utilizando o TensorFlow, implemente o modelo mostrado na figura abaixo, para classificar 2 classes diferentes.



Se nós modelarmos o classificador da figura em formato matricial, então, nós temos a seguinte equação:

$$y = f_2 \left( f_1 (XW^1 + b^1) W^2 + b^2 \right)$$

onde:

- $X$  é a matriz de entrada com dimensão  $N \times 2$ .
- $W^1$  é a matriz de coeficientes da primeira camada escondida com dimensão  $2 \times 4$ .
- $W^2$  é a matriz de coeficientes da camada de saída com dimensão  $4 \times 2$ .
- $b^1$  é o vetor de bias para a primeira camada escondida com dimensão 4.
- $b^2$  é o vetor de bias para a camada de saída com dimensão 2.
- $y$  é o vetor de saída do classificador com dimensão 2.
- $f_1(.)$  é a função logística ou sigmoid.
- $f_2(.)$  é a função softmax.
- $N$  é o número de exemplos.

Crie as 2 classes com o código abaixo.

```
from sklearn.datasets import make_circles

seed = 21

random.seed(seed)

N = 1000

X_, y_ = make_circles(n_samples=N, random_state=42, noise=0.1, factor=0.2)
```

Em seguida, faça o seguinte:

- a) Plote um gráfico mostrando as diferentes classes. (**Dica:** use cores ou marcadores diferentes para cada classe.)
- b) Crie um grafo utilizando o **GradientDescentOptimizer** para classificar os dados. (**Dica:** verifique a documentação deste otimizador no seguinte link: [https://www.tensorflow.org/api\\_docs/python/tf/compat/v1/train/GradientDescentOptimizer](https://www.tensorflow.org/api_docs/python/tf/compat/v1/train/GradientDescentOptimizer))
- c) Imprima e salve o erro (i.e., loss) e a precisão a cada 100 épocas.
- d) Treine o modelo. (**Dica:** Use 20000 épocas ou treine até que a precisão seja de 100%.)
- e) Plote um gráfico mostrando o erro e a precisão versus o número de épocas.
- f) Plote um gráfico com as fronteiras de decisão.
- g) Crie um segundo grafo utilizando o **AdamOptimizer** para classificar os dados. (**Dica:** Não se esqueça de resetar o grafo com **tf.reset\_default\_graph()**.) (**Dica:** verifique a documentação deste otimizador no seguinte link: [https://www.tensorflow.org/api\\_docs/python/tf/compat/v1/train/AdamOptimizer](https://www.tensorflow.org/api_docs/python/tf/compat/v1/train/AdamOptimizer))
- h) Imprima e salve o erro (i.e., loss) e a precisão a cada 100 épocas.
- i) Treine o modelo. (**Dica:** Use 20000 épocas ou treine até que a precisão seja de 100%.)
- j) Plote um gráfico mostrando o erro e a precisão versus o número de épocas.
- k) Plote um gráfico com as fronteiras de decisão.
- l) Baseado nos valores de erro e precisão impressos e plotados nas figuras anteriores, qual dos 2 otimizadores tem melhor performance? (**Dica:** Qual dos 2 converge mais rapidamente?)

**OBSERVAÇÃO:** Neste exercício, você vai precisar utilizar a função **tf.nn.sparse\_softmax\_cross\_entropy\_with\_logits()** para calcular o erro do modelo e assim, conseguir treiná-lo, ou seja, encontrar os pesos que minimizem o erro. A função **tf.nn.sparse\_softmax\_cross\_entropy\_with\_logits()** é equivalente a aplicar a função de ativação softmax e, em seguida, calcular a entropia cruzada, mas é mais eficiente e cuida adequadamente de casos incomuns, por exemplo, onde os logits são iguais a 0. A documentação dessa função pode ser encontrada em:

[https://www.tensorflow.org/api\\_docs/python/tf/nn/sparse\\_softmax\\_cross\\_entropy\\_with\\_logits](https://www.tensorflow.org/api_docs/python/tf/nn/sparse_softmax_cross_entropy_with_logits)

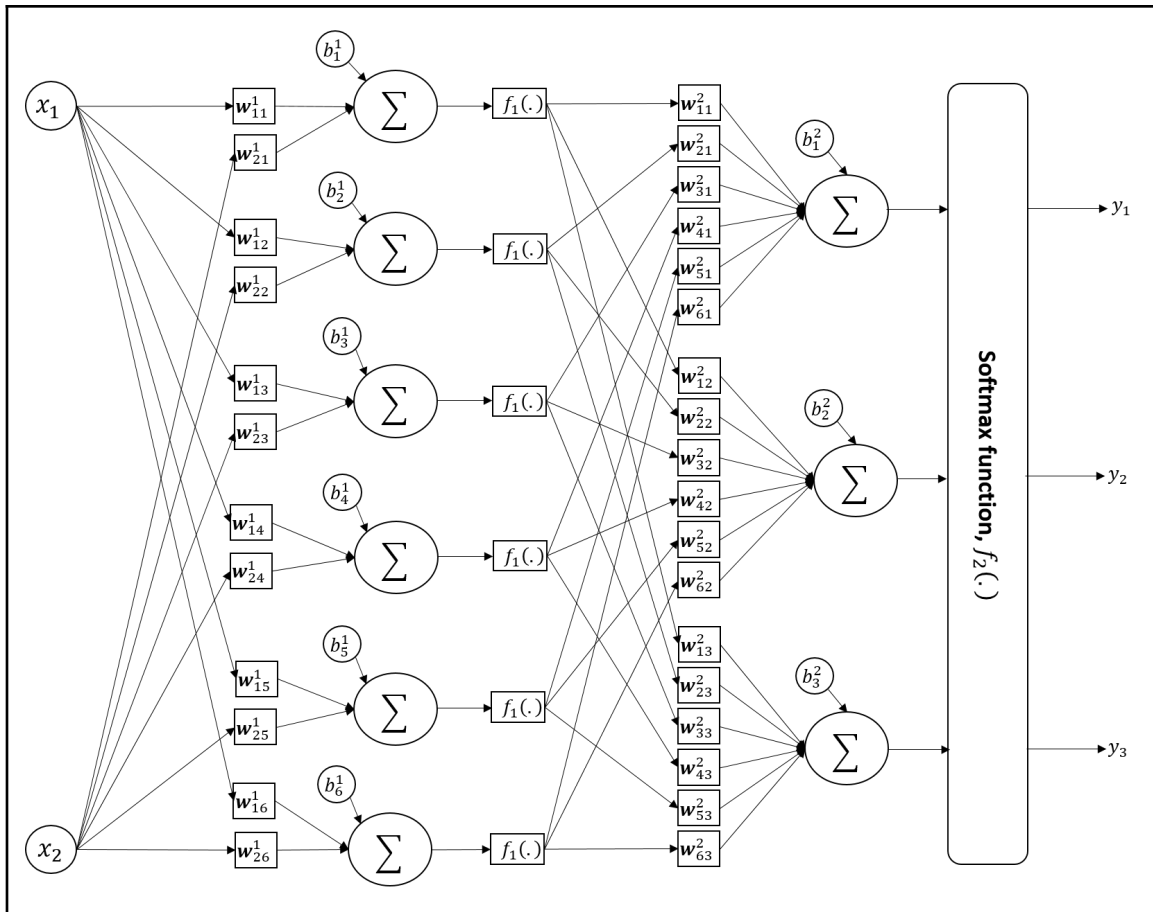
Seguem alguns exemplos de como utilizar a função:

[1] <https://towardsdatascience.com/active-learning-on-mnist-saving-on-labeling-f3971994c7ba>

[2] <https://riptutorial.com/tensorflow/example/17628/computing-costs-on-a-softmax-output-layer>

**15. Exercício sobre TensorFlow:** Utilizando o TensorFlow, implemente o modelo mostrado na figura abaixo, para classificar 3 classes diferentes.





Se nós modelarmos o classificador da figura em formato matricial, então, nós temos a seguinte equação:

$$y = f_2 \left( f_1 (XW^1 + b^1) W^2 + b^2 \right)$$

onde:

- $X$  é a matriz de entrada com dimensão  $N \times 2$ .
- $W^1$  é a matriz de coeficientes da primeira camada escondida com dimensão  $2 \times 6$ .
- $W^2$  é a matriz de coeficientes da camada de saída com dimensão  $6 \times 3$ .
- $b^1$  é o vetor de bias para a primeira camada escondida com dimensão 6.
- $b^2$  é o vetor de bias para a camada de saída com dimensão 3.
- $y$  é o vetor de saída do classificador com dimensão 3.
- $f_1(\cdot)$  é a função logística ou sigmoid.
- $f_2(\cdot)$  é a função softmax.
- $N$  é o número de exemplos.

Utilize o conjunto de dados contido no arquivo [threeMoons.csv](#) (baixe o arquivo e utilize seu conteúdo). Em seguida, faça o seguinte:

- a) Plote um gráfico mostrando as diferentes classes. (**Dica:** use cores ou marcadores diferentes para cada classe.)

- b) Crie um grafo utilizando o **GradientDescentOptimizer** para classificar os dados. (Dica: verifique a documentação deste otimizador no seguinte link: [https://www.tensorflow.org/api\\_docs/python/tf/compat/v1/train/GradientDescentOptimizer](https://www.tensorflow.org/api_docs/python/tf/compat/v1/train/GradientDescentOptimizer))
- c) Imprima e salve o erro (i.e., loss) e a acurácia a cada 100 épocas.
- d) Treine o modelo. (Dica: Treine o modelo até que sua acurácia seja de 100%.)
- e) Plote um gráfico mostrando o erro e a acurácia versus o número de épocas.
- f) Plote um gráfico com as fronteiras de decisão.
- g) Crie um segundo grafo utilizando o **AdamOptimizer** para classificar os dados. (Dica: Não se esqueça de resetar o grafo com **tf.reset\_default\_graph()**.) (Dica: verifique a documentação deste otimizador no seguinte link: [https://www.tensorflow.org/api\\_docs/python/tf/compat/v1/train/AdamOptimizer](https://www.tensorflow.org/api_docs/python/tf/compat/v1/train/AdamOptimizer))
- h) Imprima e salve o erro (i.e., loss) e a acurácia a cada 100 épocas.
- i) Treine o modelo. (Dica: Treine o modelo até que sua acurácia seja de 100%.)
- j) Plote um gráfico mostrando o erro e a acurácia versus o número de épocas.
- k) Plote um gráfico com as fronteiras de decisão.
- l) Baseado nos valores de erro e acurácia impressos e plotados nas figuras anteriores, qual dos 2 otimizadores tem melhor performance? (Dica: Qual dos 2 converge mais rapidamente?)
- m) O número de nós (ou neurônios) na camada intermediária poderia ser reduzido e ainda sim termos acurácia de 100%? Qual seria o número mínimo de nós para que ainda tivéssemos uma acurácia de 100%? Caso seja possível reduzir o número de nós, mostre isto através de um experimento.

**OBSERVAÇÃO:** Neste exercício, você vai precisar utilizar a função **tf.nn.sparse\_softmax\_cross\_entropy\_with\_logits()** para calcular o erro do modelo e assim, conseguir treiná-lo, ou seja, encontrar os pesos que minimizem o erro. A função **tf.nn.sparse\_softmax\_cross\_entropy\_with\_logits()** é equivalente a aplicar a função de ativação softmax e, em seguida, calcular a entropia cruzada, mas é mais eficiente e cuida adequadamente de casos incomuns, por exemplo, onde os logits são iguais a 0. A documentação dessa função pode ser encontrada em:

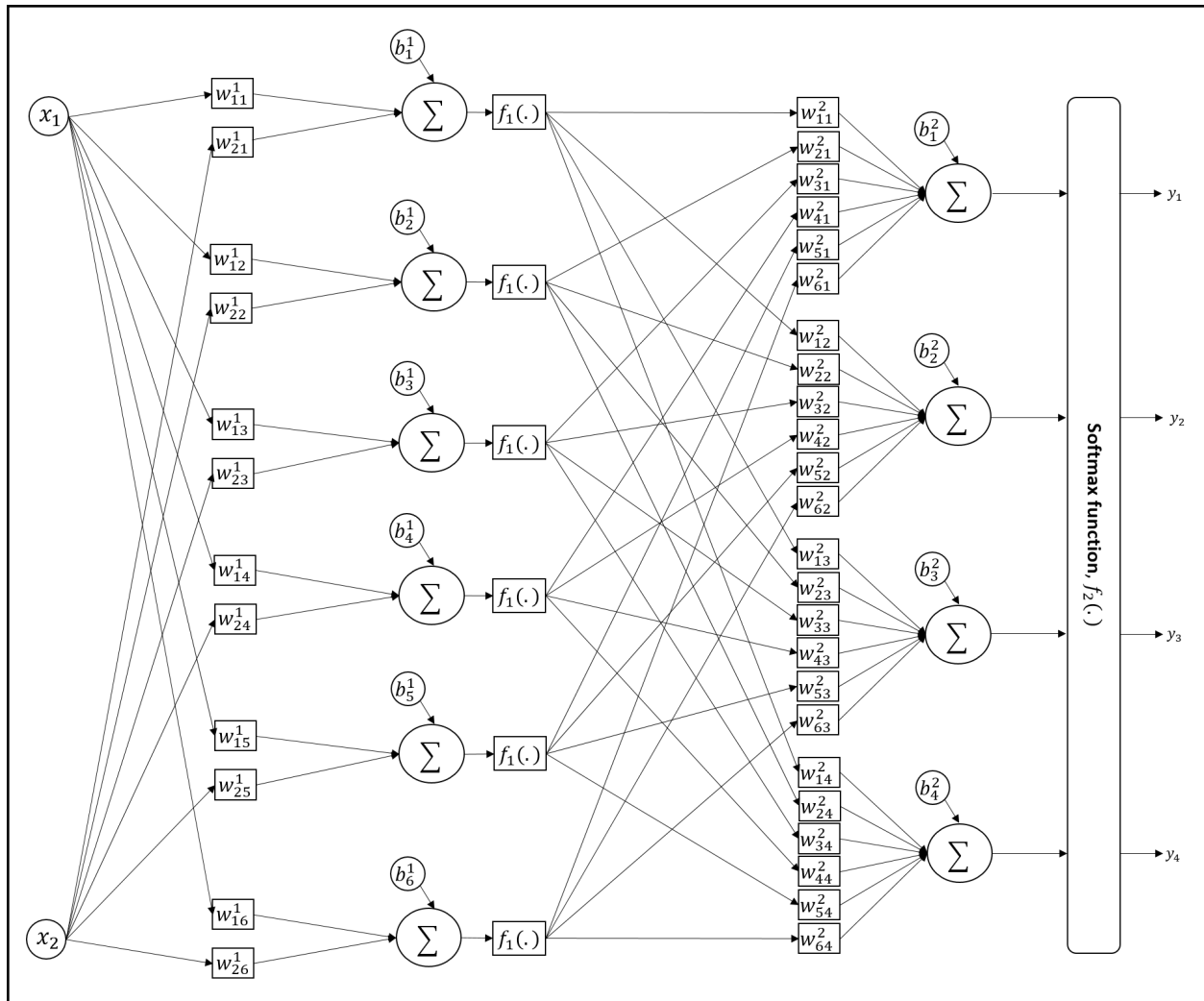
[https://www.tensorflow.org/api\\_docs/python/tf/nn/sparse\\_softmax\\_cross\\_entropy\\_with\\_logits](https://www.tensorflow.org/api_docs/python/tf/nn/sparse_softmax_cross_entropy_with_logits)

Seguem alguns exemplos de como utilizar a função:

[1] <https://towardsdatascience.com/active-learning-on-mnist-saving-on-labeling-f3971994c7ba>

[2] <https://riptutorial.com/tensorflow/example/17628/computing-costs-on-a-softmax-output-layer>

**16. Exercício sobre TensorFlow:** Utilizando o TensorFlow, implemente o modelo mostrado na figura abaixo, para classificar 4 classes diferentes.



Se nós modelarmos o classificador da figura em formato matricial, então, nós temos a seguinte equação:

$$y = f_2\left(f_1(XW^1 + b^1)W^2 + b^2\right)$$

onde:

- $X$  é a matriz de entrada com dimensão  $N \times 2$ .
- $W^1$  é a matriz de coeficientes da primeira camada escondida com dimensão  $2 \times 6$ .
- $W^2$  é a matriz de coeficientes da camada de saída com dimensão  $6 \times 4$ .
- $b^1$  é o vetor de bias para a primeira camada escondida com dimensão 6.
- $b^2$  é o vetor de bias para a camada de saída com dimensão 4.
- $y$  é o vetor de saída do classificador com dimensão 4.
- $f_1(\cdot)$  é a função logística ou sigmoid.
- $f_2(\cdot)$  é a função softmax.
- $N$  é o número de exemplos.

Utilize o conjunto de dados contido no arquivo [fourMoons.csv](#) (baixe o arquivo e utilize seu conteúdo). Em seguida, faça o seguinte:

- a) Plote um gráfico mostrando as diferentes classes. (Dica: use cores ou marcadores diferentes para cada classe.)
- b) Crie um grafo utilizando o **GradientDescentOptimizer** para classificar os dados. (Dica: verifique a documentação deste otimizador no seguinte link: [https://www.tensorflow.org/api\\_docs/python/tf/compat/v1/train/GradientDescentOptimizer](https://www.tensorflow.org/api_docs/python/tf/compat/v1/train/GradientDescentOptimizer))
- c) Imprima e salve o erro (i.e., loss) e a precisão a cada 100 épocas.
- d) Treine o modelo. (Dica: Treine o modelo até que sua precisão seja de 100%.)
- e) Plote um gráfico mostrando o erro e a precisão versus o número de épocas.
- f) Plote um gráfico com as fronteiras de decisão.
- g) Crie um segundo grafo utilizando o **AdamOptimizer** para classificar os dados. (Dica: Não se esqueça de resetar o grafo com **tf.reset\_default\_graph()**.) (Dica: verifique a documentação deste otimizador no seguinte link: [https://www.tensorflow.org/api\\_docs/python/tf/compat/v1/train/AdamOptimizer](https://www.tensorflow.org/api_docs/python/tf/compat/v1/train/AdamOptimizer))
- h) Imprima e salve o erro (i.e., loss) e a precisão a cada 100 épocas.
- i) Treine o modelo. (Dica: Treine o modelo até que sua precisão seja de 100%.)
- j) Plote um gráfico mostrando o erro e a precisão versus o número de épocas.
- k) Plote um gráfico com as fronteiras de decisão.
- l) Baseado nos valores de erro e precisão impressos e plotados nas figuras anteriores, qual dos 2 otimizadores tem melhor performance? (Dica: Qual dos 2 converge mais rapidamente?)

**OBSERVAÇÃO:** Neste exercício, você vai precisar utilizar a função **tf.nn.sparse\_softmax\_cross\_entropy\_with\_logits()** para calcular o erro do modelo e assim, conseguir treiná-lo, ou seja, encontrar os pesos que minimizem o erro. A função **tf.nn.sparse\_softmax\_cross\_entropy\_with\_logits()** é equivalente a aplicar a função de ativação softmax e, em seguida, calcular a entropia cruzada, mas é mais eficiente e cuida adequadamente de casos incomuns, por exemplo, onde os logits são iguais a 0. A documentação dessa função pode ser encontrada em:

[https://www.tensorflow.org/api\\_docs/python/tf/nn/sparse\\_softmax\\_cross\\_entropy\\_with\\_logits](https://www.tensorflow.org/api_docs/python/tf/nn/sparse_softmax_cross_entropy_with_logits)

Seguem alguns exemplos de como utilizar a função:

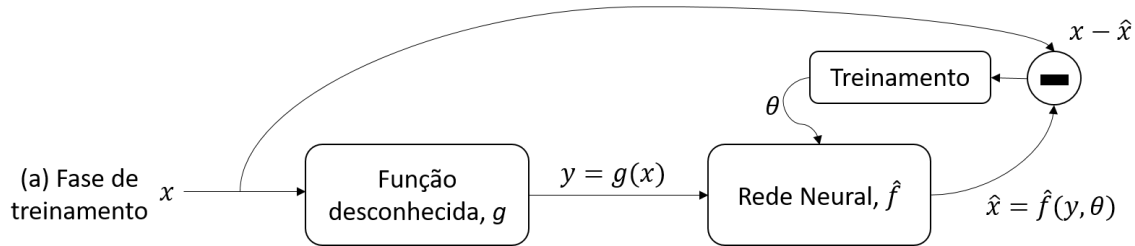
[1] <https://towardsdatascience.com/active-learning-on-mnist-saving-on-labeling-f3971994c7ba>

[2] <https://riptutorial.com/tensorflow/example/17628/computing-costs-on-a-softmax-output-layer>

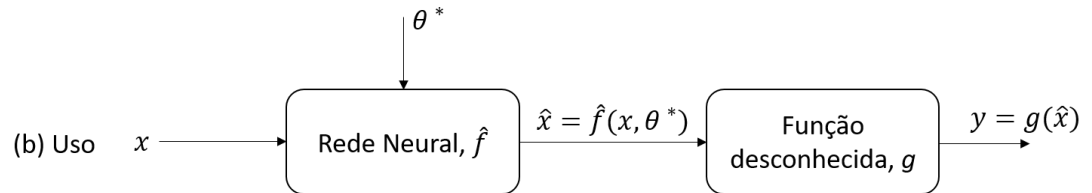
17. **Exercício sobre multi-layer perceptrons:** Neste exercício, iremos inverter uma função não-linear e usar a função invertida para linearizá-la. Em sistemas de telecomunicações, distorção não-linear pode ocorrer entre transmissor e receptor, por exemplo, devido a amplificadores não-lineares no transmissor. Esta distorção pode ser representada por uma função desconhecida  $g$  que recebe um sinal  $x$  como entrada e produz uma saída distorcida  $y = g(x)$ . A maneira convencional de desfazer a distorção é identificar um modelo parametrizado apropriado, então estimar seus parâmetros através de medições e, finalmente, criar uma função inversa com base nas estimativas. Essa abordagem está

sujeita à propagação de erros entre as três etapas. Uma abordagem alternativa é treinar um modelo de rede neural para inverter diretamente a função, sem exigir modelagem explícita ou estimativa de parâmetros. O aprendizado de máquina pode fornecer melhores resultados do que a abordagem convencional para a linearização de funções.

O procedimento geral para treinar uma rede neural para inversão de função é ilustrado na figura abaixo.



Uma função desconhecida  $g$  com entrada  $x$  é invertida usando um modelo  $\hat{f}$  treinando-o para obter  $\hat{f}(g(x), \theta^*) \approx x$ , conforme mostrado em (a). O procedimento de treinamento atualizará iterativamente os pesos  $\theta$  para reduzir gradualmente os erros de aproximação até que convirja para um valor ótimo,  $\theta^*$ .



O modelo treinado em (b) pode ser usado para linearizar a função desconhecida,  $g$ , sem ter que modelá-la explicitamente e estimar os parâmetros do modelo. Em (b), o que acontece é uma pré-distorção do sinal de entrada  $x$  o qual quando passado através da função desconhecida  $g$ , produz em sua saída um sinal linear.

Para realizar o treinamento, precisamos gerar um número  $N$  de valores  $x_n^{train}$  e passá-los através da função desconhecida,  $g$ , para medir

$$y_n^{train} = g(x_n^{train}), \text{ para } n = 1, \dots, N.$$

Então,  $y_n^{train}$  é usado como entrada para o modelo de aprendizado de máquina, enquanto  $x_n^{train}$  é a saída desejada.

**Exercício:** dada a seguinte função  $y = g(x)$ :

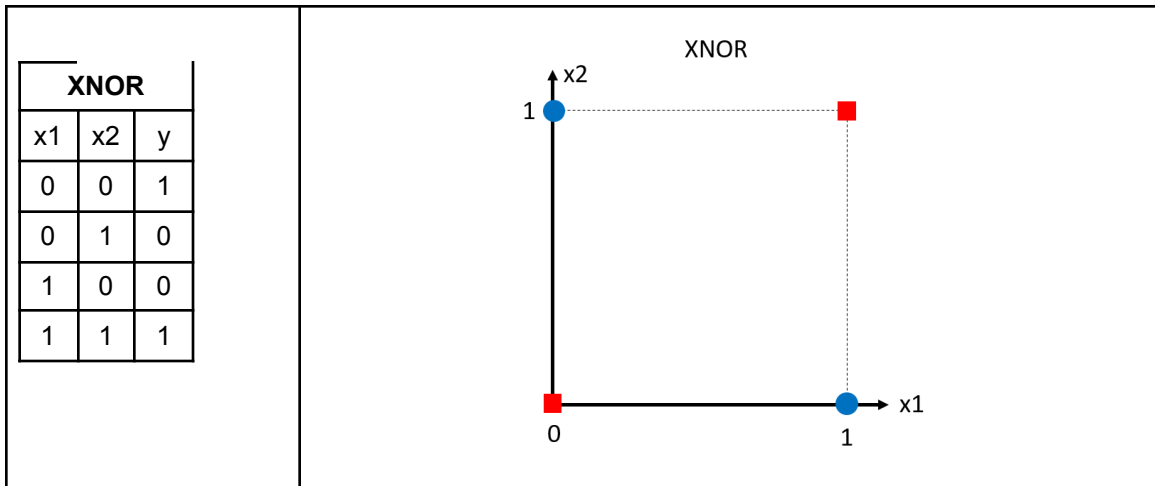
$$y = \frac{\alpha x}{\left(1 + \left(\frac{|x|}{v_{sat}}\right)^{2p}\right)^{\frac{1}{2p}}},$$

onde  $\alpha = 4$ ,  $p = 2$  e  $v_{sat} = 1$ . Use uma rede *multilayer perceptron* (MLP) para inverter a função  $g$ . Para isso, faça o seguinte

1. Plote o gráfico de  $x$  versus  $y$ .

2. Treine uma rede neural que inverta a função  $g$ .  
**(Dica 1:** use a classe **MLPRegressor** e **Grid** ou **Random Search** para encontrar o número ótimo de camadas e nós em cada camada.)  
**(Dica 2:** Comece com apenas uma camada escondida e vá aumentando gradativamente caso seja necessário para se inverter a função dada adequadamente.)
3. Após o treinamento, apresente:
  - a. Um gráfico comparando o sinal  $x_n^{train}$  com o sinal  $x_n^{pred}$  **(Dica:** use o próprio sinal de treinamento para obter a predição).
4. O modelo treinado consegue inverter, ou seja, linearizar, a função desconhecida,  $g$ ?
5. Como mostrado em (b), use o modelo treinado para pré-distorcer o sinal de entrada e, em seguida, aplique a saída do modelo à entrada da função  $g$ .
  - a. Plote um gráfico comparando
    - i. O sinal de entrada  $x$  e a saída da função  $g$ , ou seja, o sinal de saída original e distorcido.
    - ii. O sinal de entrada  $x$  e o sinal pré-distorcido  $\hat{x}$ .
    - iii. O sinal pré-distorcido  $\hat{x}$  e a saída da função  $g$  quando sua entrada é  $\hat{x}$ .
    - iv. O sinal de entrada  $x$  e o sinal de saída do sistema de pré-distorção,  $y$ , ou seja, a saída da função  $g$  quando sua entrada é  $\hat{x}$ .
  - b. O que ocorre com o sinal de saída,  $y$ ?

**18. Exercício sobre multi-layer perceptrons:** Neste exercício você irá realizar a classificação da função lógica XNOR com uma rede multi-layer perceptron (MLP). A tabela verdade e o gráfico com as classes dos elementos da tabela são apresentados abaixo. Como você deve se lembrar das listas de exercícios, classificadores lineares, entre eles o perceptron, não conseguem separar classes que não são linearmente separáveis, ou seja, no caso abaixo, uma linha reta não conseguiria criar regiões que classificariam os exemplos corretamente.



Após ler as referências abaixo, faça o seguinte

- A. Treine um **classificador** utilizando um **conjunto de perceptrons** que separe (classifique) os dados da tabela acima com precisão de 100%. **Seu classificador DEVE usar o menor número possível de perceptrons** (ou seja, de nós/neurônios) para implementar tal classificador com precisão de 100%. Gere o conjunto de treinamento com o trecho de código abaixo.

```
seed = 5
np.random.seed(seed)

N = 1000

X = np.zeros((N,2))
y = np.zeros((N,))
for i in range(N):
 x1 = np.random.randint(0, 2)
 x2 = np.random.randint(0, 2)
 X[i,:] = [x1/1.0, x2/1.0]
 y[i] = (not (x1 ^ x2))/1.0
```

(**Dica:** Utilize a classe `MLPClassifier` disponibilizada pela biblioteca `SciKit-Learn`.)

(**Dica:** Talvez você precise encontrar novos valores para os hiperparâmetros **tol** e **max\_iter**, tente, inicialmente os valores 1e-10 e 5000, respectivamente.)

(**Dica:** Lembre-se que devido a superfície de erro da rede não ser convexa, a convergência irá depender bastante da inicialização dos vetores de peso e bias, portanto, treine com vários valores de semente até que o classificador atinja precisão de 100%.)

- B. Qual é o **menor número possível de nós** (i.e., neurônios) necessários para se separar essas classes?
- C. Explique de forma sucinta como você chegou ao valor do menor número de nós (ou neurônios).

- D. Use GridSearch (i.e., a classe GridSearchCV da biblioteca SciKit-Learn) variando os parâmetros **hidden\_layer\_sizes** e **random\_state** para verificar qual é o valor ótimo para o número de nós escondidos.  
(Dica: Faça com que o GridSearch teste com o parâmetro **hidden\_layer\_sizes** variando entre 0 e 15 e o parâmetro **random\_state** variando entre 0 e 20.)
- E. Plote uma figura mostrando as fronteiras de decisão.
- F. Plote a matriz de confusão.
- G. Plote a curva ROC.
- H. Baseado na curva ROC, qual a área sob a curva?

## Referências

[1]

<https://medium.com/@jayeshbahire/the-xor-problem-in-neural-networks-50006411840b>

[2] [http://home.agh.edu.pl/~vlsi/Al/xor\\_t/en/main.htm](http://home.agh.edu.pl/~vlsi/Al/xor_t/en/main.htm)

[3] Neural network models (supervised),

[https://scikit-learn.org/stable/modules/neural\\_networks\\_supervised.html?highlight=mlp](https://scikit-learn.org/stable/modules/neural_networks_supervised.html?highlight=mlp)

[4] Perceptron,

[https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.Perceptron.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Perceptron.html)

19. **Exercício sobre multi-layer perceptron (MLP):** Dispositivos bluetooth transmitem na faixa de 2.4 GHz e utilizam *frequency hopping*, ou seja, mudam a frequência em que transmitem seus sinais, normalmente, a cada novo intervalo de transmissão. Isso causa interferência em outras tecnologias, como por exemplo Wi-Fi. Vamos supor que uma determinada faixa de frequência seja dividida em 4 canais e que dispositivos bluetooth sempre tenham o seguinte padrão de transmissão: 0, 1, 2, e 3. Ao final do intervalo do último canal, i.e., 3, o padrão se repete. Seria interessante se conseguíssemos, a partir da observação do canal atual, saber quais canais podem ser utilizados durante o próximo intervalo de transmissão para que não houvesse interferência.
- a. Você acha ser possível criar um modelo de classificação que dado o canal atual produza em sua saída a **predição do próximo canal que será utilizado**?
  - b. Se você respondeu sim à pergunta anterior, prove isto através de um experimento usando um objeto da classe **MLPClassifier** da biblioteca SciKit-Learn. Gere pelo menos 100000 pares de treinamento, ou seja, pares de vetores atributo e rótulos. O estado atual deve ser a entrada do modelo, i.e., o atributo e o próximo estado o rótulo. (Dica: Use os valores padrão para instanciar o objeto da classe **MLPClassifier**, exceto pelo parâmetro **hidden\_layer\_sizes**, o qual precisa ter seu valor ou conjunto de valores ideais encontrados. Use **GridSearch** para encontrar o valor ou conjunto de valores ideal para o parâmetro. Use os seguintes valores (,), (4), (8,4), (8,4,2) com o GridSearch.) (Dica: Use o trecho de código abaixo para gerar os conjuntos de treinamento e validação.)

```
def generateDataSet(N, numberOfClasses, cycle, numPrevStates, seed):
```

```
 np.random.seed(seed)
```



```

if(cycle == 4):
 seq = [0, 1, 2, 3]
else:
 seq = np.random.randint(0,numberOfClasses,cycle)
print('Tx pattern:',seq)

X = np.zeros((N,numOfPrevStates))
for i in range(N):
 for k in range(numOfPrevStates):
 X[i,k] = seq[(i-k) % cycle]

y = np.zeros((N,))
for i in range(N):
 y[i] = seq[(i+1) % cycle]

for i in range(20):
 print('Current state: ',X[i], ' - Next state: ', y[i])

return X, y

Seed used to reset the PN-sequence generator.
seed = 2

Number of pairs of examples.
N = 100000

Number of channels, that is also the number of classes.
numberOfClasses = 4

Number of states before repeating the sequence.
cycle = 4

Define the number of previous states to be used as attributes.
numOfPrevStates = 1

X, y = generateDataSet(N, numberOfClasses, cycle, numOfPrevStates, seed)

Split array into random train and test subsets.
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=seed)

```

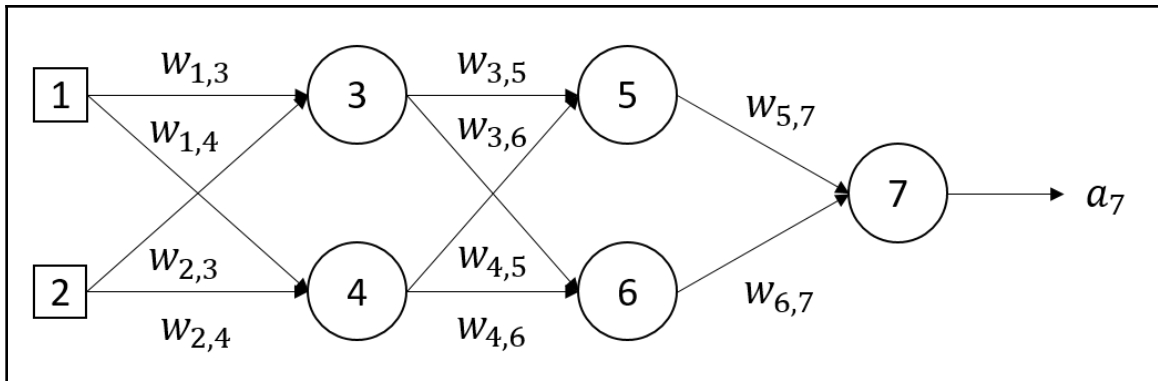
- c. Plote a matriz de confusão deste modelo utilizando o conjunto de validação.
- d. Qual a acurácia obtida?
- e. O que aconteceria com seu modelo treinado anteriormente se agora o padrão de transmissão fosse 0, 3, 1, 0, 2, 3, 2, 3, 0, e 3. Após o último canal, a sequência se repete. Use a função apresentada no trecho de código acima para gerar um

conjunto de exemplos com a nova sequência de transmissão, para isso, altere o valor do parâmetro **cycle** de 4 para 10 e execute a função novamente. Não se esqueça de dividir o novo conjunto em conjuntos de treinamento e de validação. Na sequência, apresente ao modelo o novo conjunto de treinamento. Plote a matriz de confusão e a imprima a acurácia do modelo.

- f. O modelo consegue prever qual será o próximo estado? Ou seja, ele consegue dizer qual será o próximo canal a ser utilizado no próximo intervalo de transmissão?
- g. Com a nova sequência de canais, você acha ser possível treinar um outro modelo de classificação, usando um objeto da classe **MLPClassifier** novamente, que **dado o canal (i.e., estado) atual e/ou canais anteriores**, i.e., estados anteriores, produza em sua saída a **predição do próximo canal que será utilizado**? Se necessário, gere novos conjuntos de treinamento e validação considerando o estado atual e/ou os estados anteriores como atributos de entrada para o modelo de classificação. Não se esqueça que o parâmetro **cycle** da função **generateDataSet** deve ser feito igual a 10. (**Dica:** O parâmetro **numOfPrevStates** da função **generateDataSet** altera a quantidade de estados atual e anteriores usados para se criar um vetor de atributos de entrada para o modelo de classificação.) (**Dica:** O vetor de atributos formado através da combinação do estado atual com estados anteriores pode ser entendido como um estado.) (**Dica:** Caso você se depare com a seguinte mensagem durante o treinamento: “**ConvergenceWarning: Stochastic Optimizer: Maximum iterations (200) reached and the optimization hasn’t converged yet**”, aumente o valor do parâmetro **max\_iter** de seu valor padrão igual a 200 para 1000.)
- h. Plote a matriz de confusão deste novo modelo utilizando o novo conjunto de validação.
- i. Qual a acurácia obtida?
- j. Foi possível se predizer o próximo canal que será utilizado usando a sequência 0, 3, 1, 0, 2, 3, 2, 3, 0, e 3? Caso não, explique o motivo do modelo não ser capaz de predizer o próximo canal.

20. Dada a rede MLP da figura abaixo, encontre

- a. A derivada parcial do erro de saída,  $Loss_7 = (y_7 - a_7)^2$ , em relação ao peso  $w_{1,3}$ , i.e.,  $\frac{\partial Loss_7}{\partial w_{1,3}}$ .
- b. Uma forma geral para a atualização dos pesos da primeira camada escondida, ou seja, para os pesos  $w_{1,3}$ ,  $w_{1,4}$ ,  $w_{2,3}$  e  $w_{2,4}$ .  
(**Dica:** Siga a análise feita no documento: “Aprendizado em Redes Neurais com Múltiplas Camadas”).



21. Neste exercício, você irá investigar o uso de MLPs empregando funções de ativação sigmóide para obter mapeamentos um-para-um, conforme descrito na sequência:

- $f(x) = 1/x$ ,  $1 \leq x \leq 100$ .
- $f(x) = \log_{10}(x)$ ,  $1 \leq x \leq 10$ .
- $f(x) = \exp(-x)$ ,  $1 \leq x \leq 10$ .
- $f(x) = \sin(x)$ ,  $0 \leq x \leq \pi/2$ .

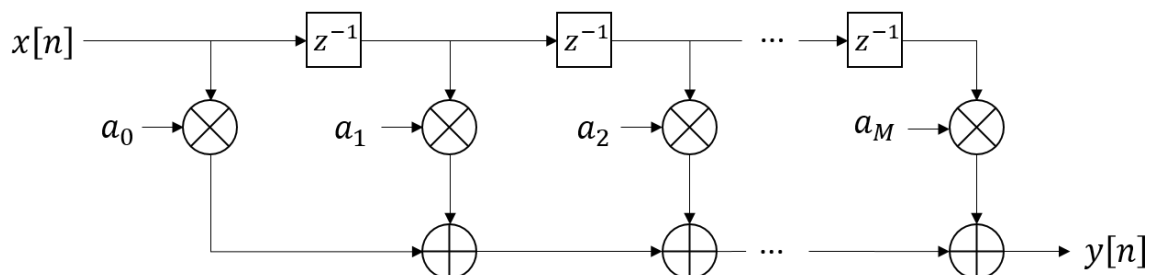
Para cada mapeamento, faça o seguinte:

- Divida os dados em dois conjuntos, um para treinamento e outro para teste.
- Use o conjunto de treinamento para treinar a rede, **com uma única camada escondida**.
- Avalie a precisão do cálculo da rede usando os dados de teste.
- Apresente figuras comparando a predição feita com cada uma das MLPs treinadas com os dados originais.
- O que você pode concluir após realizar estes vários mapeamentos com uma MLP com uma camada escondida?

Use uma única camada escondida, mas com um número variável de nós escondidos. Investigue como o desempenho da rede é afetado pela variação do tamanho da camada escondida.

(Dica: Use grid search com alguns valores para encontrar o número ótimo de nós na camada escondida)

22. **Treinando um modelo de regressão (rede neural) para se comportar como um filtro FIR:** Um filtro FIR (do inglês, Finite Impulse Response) de ordem  $M$ , tem a seguinte estrutura



Essa estrutura pode ser representada matematicamente através da seguinte equação

$$y[n] = a_0 x[n] + a_1 x[n - 1] + a_2 x[n - 2] + \dots + a_M x[n - M] = \sum_{k=0}^M a_k x[n - k],$$

onde  $M$  é a ordem do filtro,  $x$  é o sinal de entrada do filtro,  $y$  é a saída do filtro, ou seja, o sinal filtrado e  $a_k$ ,  $\forall k$  são os coeficientes do filtro.

Perceba que esta é uma equação linear, onde temos uma combinação linear dos atributos, neste caso versões atrasadas do sinal de entrada  $x[n]$ , em relação aos pesos, que são os coeficientes do filtro.

Se a ordem do filtro, ou seja, o parâmetro  $M$ , já estiver definido previamente, só é necessário encontrar seus coeficientes.

Assim, podemos aplicar algoritmos de regressão linear, incluindo redes neurais, para encontrar esses coeficientes.

Agora, vamos supor que temos um filtro operando em um sistema de telecomunicações e que queremos replicar este filtro para uso em outro sistema. Porém, nós não conhecemos o projeto do filtro. Ou seja, seu tipo (e.g., passa-baixas, passa-altas, etc.), sua ordem, frequência de corte, etc.

Como podemos projetar um filtro que se aproxime das características do filtro desejado?

Uma possível abordagem é injetar ruído branco Gaussiano (i.e., média igual a zero e variância unitária) em sua entrada e usar as entradas e a saídas resultantes para treinar uma rede neural para regressão.

Use o seguinte exemplo como base para resolver este exercício:  
[https://colab.research.google.com/github/zz4fap/tp555-ml/blob/main/misc/Exercicio\\_filtro\\_fir\\_mlp.ipynb](https://colab.research.google.com/github/zz4fap/tp555-ml/blob/main/misc/Exercicio_filtro_fir_mlp.ipynb)

Após estudar o notebook de exemplo, faça o seguinte:

- Gere 1000 amostras retiradas de uma distribuição Gaussiana normal padrão.
- Passe essas amostras através do filtro FIR dado no notebook de exemplo
  - Estude o código presente no notebook de exemplo para entender como fazer isso.
  - A ordem do filtro FIR do notebook de exemplo é igual a 45.
- Note que a equação acima, que representa o filtro FIR, é a equação de um modelo linear
  - A função **createDataset** cria uma matriz de atributos de entrada de acordo com a equação que representa o filtro FIR, ou seja, gera os atributos da função hipótese desejada. Portanto, use esta função para gerar a matriz que será usada durante o treinamento do modelo de regressão.

- Cada linha da matriz de atributos retornada pela função **`createDataset`** corresponde a um elemento da array de saída do filtro FIR, ou seja, o valor esperado (rótulo) para a n-ésima linha da matriz de atributos corresponde ao n-ésimo elemento do vetor de saída do filtro.
- Portanto, temos uma matriz de atributos com dimensões  $N \times M + 1$  e um vetor de rótulos (i.e., valores esperados) com dimensões  $N \times 1$  que devem ser usados para o treinamento do modelo de regressão.
- Use um objeto da classe **`MLPRegressor`** para treinar o modelo de regressão.
- Após o treinamento do modelo de regressão, faça:
  - Calcule e imprima o valor do erro quadrático médio (use todo o conjunto de dados).
  - Imprima e compare os pesos do modelo com os coeficientes do filtro FIR. Eles são próximos?
  - Passe o sinal de teste (sinal com a soma de dois sinais cossenoidais gerados no notebook de exemplo) através do modelo de regressão linear (use a função **`createDataset`** para converter o sinal de teste em uma matriz e o método **`predict`** do modelo de regressão para filtrar o sinal de teste).
  - Plote a FFT do sinal de saída do modelo de regressão.
  - Compare a FFT do sinal de saída do modelo de regressão com a FFT do sinal de saída do filtro FIR. Os resultados são similares?

## Referências

[1] 'FIR filter', <https://scipy-cookbook.readthedocs.io/items/FIRFilter.html>

[2] 'Filters', <https://pysdr.org/content/filters.html>

## 23. Exercício sobre Autoencoders para aprendizado de sistemas de comunicação

**fim-a-fim:** Neste exercício, usaremos um tipo de rede neural conhecida como autoencoder. Autoencoders são modelos de aprendizado não-supervisionado. Eles são um tipo de rede neural que aprende a copiar sua entrada para sua saída. Seu objetivo é aprender uma representação para um conjunto de dados, para redução ou aumento de dimensionalidade. Essas redes neurais são compostas por 2 partes: (i) **codificador**, que encontra representações eficientes para as entradas com menor ou maior dimensionalidade e (ii) **decodificador** que reconstrói a entrada.

Após esta breve introdução sobre o que são autoencoders, vamos utilizá-los para implementar um sistema de comunicações fim-a-fim. Este sistema, do lado do transmissor, irá receber como entrada uma mensagem binária composta por  $k$  bits, irá convertê-las na representação one-hot encoding (ou seja, teremos  $M = 2^k$  possíveis mensagens de  $k$  bits), encontrar uma representação robusta de  $n$  símbolos reais da mensagem de entrada. Esta representação robusta com  $n$  símbolos trafega por um canal ruidoso e, do lado do receptor, ela é decodificada de tal forma que a mensagem original possa ser recuperada com a menor probabilidade de erro possível.

O transmissor é composto por duas camadas densas (`tf.keras.layers.Dense`) com  $M$  e  $n$  nós respectivamente e recebe como entrada as representações one-hot encoding (`tf.one_hot`) das mensagens de  $k$  bits. A primeira camada utiliza ativação do

tipo *relu* e a segunda não utiliza nenhum tipo de ativação. Finalmente, para evitar que o transmissor aprenda valores de saída desnecessariamente grandes e se torne numericamente instável, normalizamos a potência média de todas as saídas do transmissor no mini-batch para um valor igual a 1, conforme mostrado abaixo.

```
x = tx/tf.sqrt(tf.reduce_mean(tf.square(tx)))
```

(OBS.: Devemos criar uma *placeholder* (tf.placeholder) para que possamos alterar o tamanho dos *mini-batches* em tempo de execução.)

O canal é implementado como um canal de ruído gaussiano branco aditivo (AWGN) o qual adiciona ruído gaussiano com média igual a zero e potência de ruído igual a  $\sigma^2$  aos símbolos transmitidos, conforme mostrado abaixo.

```
noise = tf.random.normal(shape=tf.shape(x), stddev=noise_std)
y = x + noise
```

(OBS.: Devemos criar uma *placeholder* (tf.placeholder) para que possamos alterar o desvio padrão do ruído em tempo de execução.)

O receptor é composto por duas camadas densas (tf.keras.layers.Dense) ambas com M nós. A primeira camada utiliza ativação do tipo *relu* e a segunda utiliza ativação do tipo *softmax*. A saída do receptor é um vetor de probabilidades que atribui uma probabilidade a cada uma das mensagens possíveis. Finalmente, a mensagem original é recuperada selecionando a saída com maior probabilidade.

O autoencoder resultante pode ser treinado fim-a-fim usando o gradiente descendente estocástico (SGD).

(OBS.: Devemos criar uma *placeholder* (tf.placeholder) para que possamos alterar o passo de aprendizagem em tempo de execução.)

#### Dicas de implementação:

- O mini-batch de mensagens que queremos transmitir são retirados de uma distribuição aleatória uniforme, conforme mostrado abaixo.

```
s = tf.random.uniform(shape=[batch_size], minval=0, maxval=M, dtype=tf.int32)
```

- E para alimentá-los com eficiência à primeira camada densa do transmissor, nós as transformamos nos chamados vetores (ou tensor) one-hot. Este vetor agora contém *batch\_size* vetores de comprimento M, onde apenas um elemento é definido como 1.0, enquanto todos os outros elementos são 0.0. Este tipo de vetor é comumente usado para tarefas de classificação.

```
s_one_hot = tf.one_hot(s, depth=M)
```

- Precisamos definir também uma função de custo que calcule o desempenho atual do modelo comparando a entrada com a saída. Usamos uma função de custo de entropia cruzada padrão que inerentemente ativa os logits com “softmax” e aceita rótulos esparsos, conforme mostrado abaixo.

```
cross_entropy = tf.losses.sparse_softmax_cross_entropy(labels=s, logits=s_hat)
```

- Para calcularmos a taxa média de erro de mensagem (ou de bloco) do mini-batch usamos decisão rígida com base no elemento com a maior probabilidade (argmax), conforme mostrado no trecho de código abaixo.

```
correct_predictions = tf.equal(tf.argmax(tf.nn.softmax(s_hat), axis=1, output_type=tf.int32), s)
accuracy = tf.reduce_mean(tf.cast(correct_predictions, dtype=tf.float32))
bler = 1.0 - accuracy
```

- Também precisamos definir um algoritmo otimizador que atualize os pesos de nosso autoencoder de acordo com a perda atual e o gradiente do mini-lote. Escolhemos o otimizador Adam para minimizar nossa função de custo e usar um *placeholder* para o passo de aprendizagem para poder ajustar este hiperparâmetro durante o treinamento.

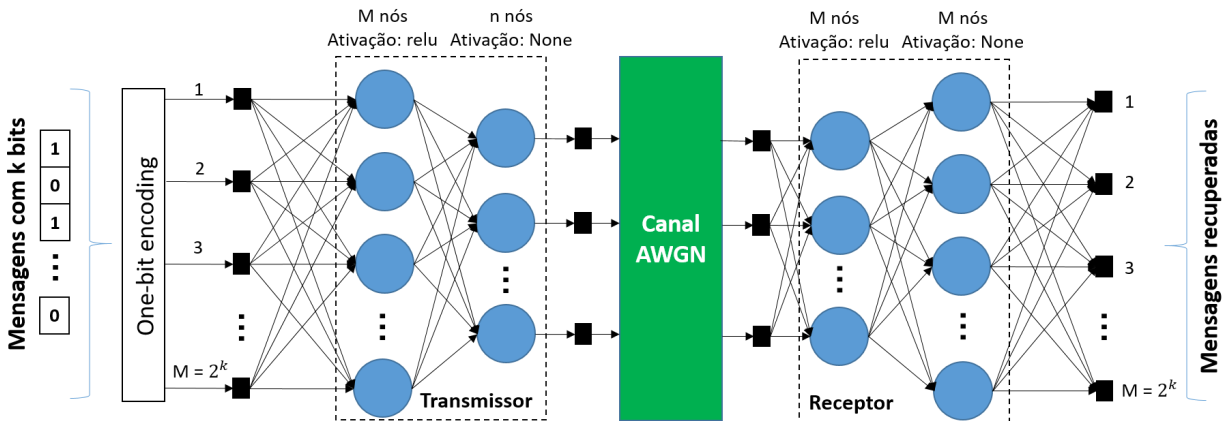
```
lr = tf.placeholder(dtype=tf.float32, shape=[])
train_op = tf.train.AdamOptimizer(learning_rate=lr).minimize(cross_entropy)
```

- Antes de iniciarmos o treinamento, precisamos definir a relação sinal-ruído (SNR), para que possamos treinar o modelo para um valor de SNR desejado. Esta função simplesmente calcula o desvio padrão do ruído para um determinado valor de SNR (enquanto a potência do sinal é normalizada para 1.0).

```
def EbNo2Sigma(ebnodb):
 ebno = 10**(ebnodb/10)
 bits_per_complex_symbol = k/(n/2)
 return 1.0/np.sqrt(bits_per_complex_symbol*ebno)
```

- Durante todas as épocas de treinamento, devemos definir o SNR como 7.0 dB, pois foi percebido que treinar um autoencoder em uma taxa de erro de bloco (BLER) de cerca de 0.01 leva a uma generalização rápida por parte do modelo. (**Dica:** para o treinamento, use um mini-batch com 1000 exemplos, passo de aprendizagem igual a 0.0001 e cerca de 100000 iterações, mas não se prenda a estes valores, teste e verifique se existem valores melhores.)

A figura abaixo mostra o sistema fim-a-fim que teremos ao final da implementação.



Agora, depois de termos implementado o sistema fim-a-fim com autoencoder, vamos verificar seu desempenho plotando sua BLER vs SNR em uma faixa de valores de SNR para  $k = 8$  e  $n = 8$ . Quando o número de bits,  $k$ , e o número de usos reais do canal são iguais a 8, temos que será gerados  $n/2=4$  símbolos complexos onde cada um destes símbolos carrega  $k/(n/2)=2$  bits de informação, o que equivale à modulação QPSK. Portanto, precisamos executar uma simulação de Monte Carlo para obter a BLER para cada valor de SNR. Neste exercício, iremos variar a SNR de 0 a 14 dB em passos de 1 dB executando 10 mini-batches de 100.000 mensagens para cada valor de SNR. Em seguida, devemos comparar, através de uma figura, a BLER obtida com o modelo com a BLER de um sistema convencional de modulação, no caso, QPSK, a qual é dada através do seguinte vetor:

```
BLER_QPSK = np.array([4.80998e-01, 3.71098e-01, 2.63797e-01, 1.68919e-01,
9.54540e-02, 4.68860e-02, 1.89070e-02, 6.11700e-03, 1.58300e-03, 2.91000e-04,
3.30000e-05, 2.00000e-06])
```

Se o treinamento do modelo foi bem sucedido, nós iremos verificar que a BLER do autoencoder é menor do que o da modulação QPSK em toda a faixa de valores de SNR. Se você estiver interessado nas origens desse ganho, dê uma olhada em [1] e [2].

#### Referências:

- [1] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," IEEE Transactions on Cognitive Communications and Networking, vol. 3, no. 4, pp. 563-575, Oct. 2017.
- [2] S. Dörner, S. Cammerer, J. Hoydis, and S. ten Brink, "Deep Learning-based Communication Over the Air," IEEE J. Sel. Topics Signal Process., vol. 12, no. 1, pp. 132-143, Feb. 2018.

**24. Exercício sobre multi-layer perceptron (MLP):** Neste exercício, você irá utilizar uma rede MLP para realizar a transformada rápida de Fourier (Fast Fourier Transform, FFT) de sinais. Entretanto, você irá treinar a rede MLP para tais transformações apenas utilizando amostras ruidosas e suas respectivas transformadas de Fourier. Agora, faça o seguinte:

1. Iremos utilizar o sinal criado com o trecho de código abaixo para verificar se a rede MLP aprendeu como realizar a transformada de Fourier. Execute o código e



verifique que o sinal é composto pela soma de 3 sinais senoidais com frequências iguais a 2, 4 e 6 Hz, respectivamente.

```
seed = 42
np.random.seed(seed)

seed = 42
np.random.seed(seed)

N = 64

k = 2

n = np.arange(0,N) - (N/2)

y = np.sin(2*np.pi*(k/N)*n) + 2*np.sin(2*np.pi*2*(k/N)*n) + 4*np.sin(2*np.pi*3*(k/N)*n)

plt.plot(n,y)
plt.grid()
plt.xlabel('n (sample number)', fontsize=14)
plt.ylabel('y', fontsize=14)
plt.show()
```

2. A transformada de fourier do sinal y é encontrada com o trecho de código abaixo.

```
Y = np.fft.fft(y)
```

3. O trecho de código abaixo plota um gráfico com o valor absoluto do sinal Y, ou seja, a transformada de Fourier do sinal y.

```
Y = Y.reshape(N,)
Y_abs = np.abs(Y)
freqs = np.fft.fftfreq(N, d=(1.0/(N*1.0)))

plt.plot(freqs[0:32], Y_abs[0:32], '--.b')
plt.plot(freqs[32:N], Y_abs[32:N], '--.b')
plt.xlim(-10,10)
plt.xticks(range(-10,10,2))
plt.xlabel('Frequency', fontsize=14)
plt.ylabel('$|y|$', fontsize=14)
plt.grid()
plt.show()
```

4. Gere 10000 pares de vetores atributo e rótulos. Os vetores atributo são gerados a partir de uma distribuição Gaussiana Normal com média igual a zero e desvio padrão igual a 10. Cada vetor atributo deve possuir 128 elementos, criados a partir da intercalação entre as partes real e imaginária dos valores complexos retirados da distribuição Gaussiana. Use o trecho de código abaixo para gerar o conjunto de pares de vetores atributo e rótulos. OBS.: Note que o conjunto já é

dividido em conjuntos de treinamento e validação, portanto, a divisão não é necessária.

```
M = 10000

X = np.zeros((M,2*N))
y = np.zeros((M,2*N))

for i in range(M):
 s = np.random.normal(loc=0.0, scale=10.0, size=(N,))
 S = np.fft.fft(s)
 X[i,0::2] = s.real
 X[i,1::2] = s.imag
 y[i,0::2] = S.real
 y[i,1::2] = S.imag

X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=seed)
```

5. Usando GridSearch, encontre o conjunto de hiperparâmetros ideais para que a rede MLP aprenda a realizar a transformada de Fourier. Use o seguinte conjunto de valores para verificar quais deles são ideais para o aprendizado da rede MLP. (**Dica.**: Talvez seja necessário ajustar o valor do hiperparâmetro **max\_iter** da classe **MLPRegressor**, use inicialmente valor 1000, caso não resolva, aumente o valor).

```
'Hidden_layer_sizes' : [(N,),(N,N),(N,N,N)],
'learning_rate' : ['adaptive','invscaling','constant'],
'activation':['relu','logistic'], 'batch_size' : [64, 128, 256]
```

6. Na sequência, usando os melhores hiperparâmetros, compare a saída da rede MLP com a saída da FFT da biblioteca numpy quando usamos o sinal y (aquele com a soma de três sinais senoidais) como entrada para ambos. Plote uma figura comparando ambos os resultados.
7. Após observar o resultado do item anterior, o que você pode concluir?