

## Lista de exercícios - Machine Learning

Arthur de Andrade Almeida e Wallace Nascimento Melo  
Recife, 2020

**Problema 1** Como um classificador não sequencial pode ser modificado para identificar padrões sequenciais?

*Resposta* Circuitos digitais podem ser divididos em dois grupos principais: Circuitos combinacionais e circuitos sequenciais. A principal diferença é que esse último possui algum tipo de dispositivo de memória, a qual são utilizados para realizar operações com estados anteriores do circuito. Como exemplo, podemos criar um registrador com  $k$  *flip-flops* que armazena valores de uma variável de um instante  $x(n)$  até um instante  $x(n - k)$  e entrega esses valores para um sistema puramente combinacional que irá processar a relação (reconhecer os padrões) entre essas variáveis. No caso de redes neurais artificiais, ao invés de dispositivos de memória e portas lógicas, teremos memórias de valores de reais e os nós (ou neurônios) em cada camada, respectivamente. No entanto, a maior diferença reside no fato do algoritmo de treinamento da rede, que pode ser *backpropagation* ou variação que leva a sequencialidade em consideração, como o *backpropagation through time (bptt)*.

**Problema 2** Descreva o que caracteriza os seguintes problemas na área de aprendizagem de máquina: classificação, regressão, clustering.

*Resposta* O aprendizado de máquina nos possibilita lidar com **tarefas** que são muito difíceis de resolver com programas fixos escritos e projetados por seres humanos. Classificação, regressão e *clustering* são algumas das tarefas mais comuns de aprendizagem de máquina.

**Classificação:** Neste tipo de tarefa supervisionada, o programa deverá especificar qual das  $k$  categorias os exemplos de entrada pertencem. Para resolver esta tarefa, o algoritmo de aprendizagem produz uma função  $f : \mathbb{R}^n \rightarrow 1, \dots, k$ . Quando  $y = f(\vec{x})$ , o modelo atribui a entrada  $\vec{x}$  algum dos grupos indicados no conjunto imagem. Um exemplo de classificação é reconhecimento de objetos, onde a entrada é uma imagem (cada pixel é uma *feature*), e a saída é um código numérico indicando o objeto na imagem.

**Regressão:** Nesta tarefa, dada a entrada o programa deverá prever um valor numérico. Para resolver uma tarefa de regressão, o algoritmo de aprendizagem é utilizado para gerar uma função  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . Este tipo de tarefa é bem próximo de uma classificação (ambas são supervisionadas), embora na classificação a saída seja um valor discreto. Um exemplo de regressão pode ser prever valores de venda de algum produto.

**Clustering:** Diferentemente das tarefas citadas acima, o *clustering* é uma tarefa sem supervisão, ou seja, inicialmente os dados não possuem rótulos. O principal objetivo de uma tarefa de *clustering* é interpretar os dados de entrada e procurar agrupá-los de acordo com as similaridades de suas características.

**Problema 3** Em reconhecimento de padrões, o que é um problema linearmente separável?

*Resposta* Um problema linearmente separável, é aquele que a sua representação cuja representação de seu conjuntos de dados possa ser separada por uma reta, plano ou hiperplano. Alguns problemas mais conhecidos, como o do OR, AND e NOT, podem ser resolvidos por métodos lineares, sendo alguns deles Perceptrons ou Classificadores Bayesianos, porém outros como o problema do XOR, não podem ter seus grupos separados por linearização, e por isso não pode ser alcançados por um único perceptron por exemplo, e necessita de arquiteturas mais avançadas de redes artificiais.

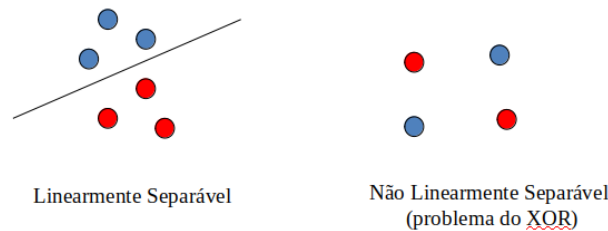


Figura 1: Exemplificação de problema linearizável e não-linearizável.

**Problema 4** Como um classificador linear pode ser modificado para identificar padrões não lineares?

*Resposta* No caso de um problema não ser linearmente separável, pode-se projetar os dados em um espaço de maior dimensão no quais esses se tornem linearmente separáveis. Após encontrar um hiperplano em que seja possível separar os dados nesse novo espaço, basta utilizar uma função de transformação não linear, e com vetor de imagem criado, pode-se classificar linearmente. Essa solução é cabível para o problema não-linearizável do XOR, e permite que esse seja classificado de forma linear.

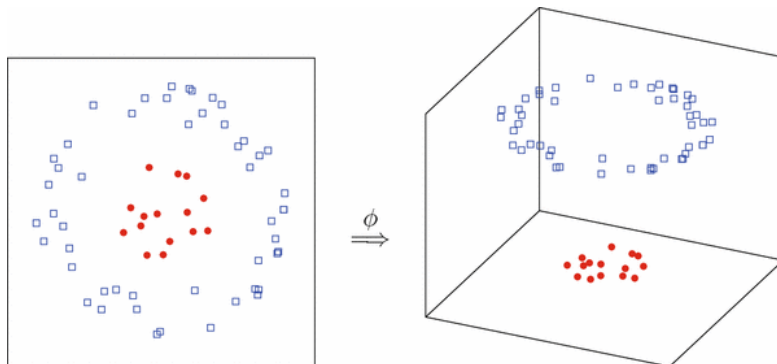


Figura 2: Transformação para dimensão Superior.

**Problema 5** Considerando o problema do "XOR"(Ver lista do monitor). Quais classificadores poderiam ser utilizados?

- ☐ Perceptron
- ☒ Rede Neural MLP
- ☒ Rede Neural RBF
- ☐ Classificador Bayesiano
- ☒ KNN

**Problema 6** Considerando a base de dados do Iris, implemente uma rede MLP com 4 entradas (uma para cada atributo), uma camada intermediária com ativação Relu, uma camada de saída com ativação sigmoide e três saídas (uma para cada classe da base de dados). O número de nodos na camadas intermediária deve ser ajustado para garantir uma taxa de acerto superior a 90% sem overfitting. No treinamento deve ser utilizado validação cruzada k-fold com k=5. O treinamento deve ser feito no modo batch (em grupos de 5 amostras). O critério de parada deve ser 1000 épocas ou divergência ou estabilização nas curvas de erro do conjunto de treinamento e de validação por mais que 5 épocas.

*Resposta* **P5.py**

**Problema 7** Considerando a base de dados do Iris, calcule os vetores de média, matrizes de covariância e probabilidades a priori para cada uma das classes da base de dados. Represente os vetores e matrizes com a mesma sintaxe utilizada no Scilab. (notação vetor coluna, "["para delimitar o inicio

do vetor ou matriz, "]" para delimitar o fim do vetor ou matrize, ";" para delimitar as linhas e "," para as colunas.)

*Resposta* **P6.py**

**Problema 8** Considerando a base de dados "wine", implemente um classificador para determinar a que classe um determinado vinho pertence. O classificador deve apresentar uma saída por classe e deve ser um das seguintes opções: KNN, RBF, Bayesiano, LS.

*Resposta* **P7.py**

**Problema 9** Considerando a base de dados "Pulso", implemente um classificador para detectar a ocorrência de pulsos consecutivos. A base de dados "Pulso" consiste de uma matriz de 400 linhas por 2001 colunas. Cada linha da base de dados corresponde a uma janela de dimensão 2000 concatenada com um rótulo binário indicando ou não a presença do sinal a ser detectado. O classificador deve ser um das seguintes opções: KNN, MLP, RBF, CNN, Bayesiano, LS.

*Resposta* **P8.py**