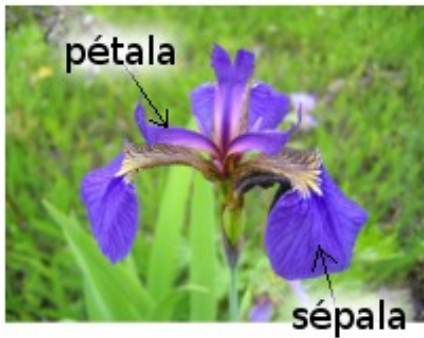


Exercícios – RNA – Slide 03

Um conjunto de dados famoso na classificação é composto de dados da flor Iris e suas três espécies:



Iris Setosa



Iris Virginica



Iris Versicolor

Entre no diretório onde estão os arquivos “iris.data” e “iris.names”:

```
>>> cd diretório
```

Carregue o arquivo “iris.data” com a instrução:

```
>>> X = load('iris.data');
```

As quatro primeiras colunas são as entradas e a última coluna é a definição da espécie (classe) da Iris. Leia o conteúdo do arquivo “iris.names” para compreender melhor.

Notas:

Para saber quais são as linhas que lhe fornecem a Iris da classe 0 (zero), você deve fazer:

```
>>> find( X(:, 5) == 1 )
```

Pode-se obter somente as linhas desejadas – por ex.: linhas 1, 2, 3 e 4 – com a instrução:

```
>>> X( [1 2 3 4], : )
```

Obtenha então somente as linhas da Iris Setosa e da Versicolor.

Agora, separe essa matriz em duas variáveis, sendo:

```
X ← entradas
```

```
y ← classes
```

Verifique qual é Variância (*octave: var*) dos valores das entradas X. O que elas significam?

Verifique também qual é a Covariância (*octave: cov*) das entradas X. O que elas significam?

Crie um vetor linha com a coluna que possui a maior variância de sua matriz:

```
X1 ← X(:, NUMERO_DA_COLUNA)
```

Agora, visualize os valores de X1 e de y:

```
>>> disp( [ X1 y ] )
```

Plote um gráfico mostrando os valores de entrada nas abscisas e o valor de saída nas ordenadas.

Como classificar a Iris como Setosa ou Versicolor usando somente esses valores de entrada?

Com essas entradas, é possível separar os três tipos de Iris? Por quê? Qual é a taxa de acerto?

Continuemos o exercício utilizando todas as entradas e todas as espécies de flores...

Plote gráficos em duas dimensões com os seguintes atributos das flores:

comprimento da sépala *por* largura da sépala
comprimento da sépala *por* comprimento da pétala
comprimento da sépala *por* largura da pétala

largura da sépala *por* comprimento da pétala
largura da sépala *por* largura da pétala

comprimento da pétala *por* largura da pétala

Quais informações tornam-se mais claras com essas visualizações?

Faça a estandardização para cada atributo dos padrões de entrada.

Verifique a variância e a covariância desses dados. Qual a diferença dos valores antigos?

Formule uma função que identifique qual é a distância entre os padrões de entrada.

Distância Euclidiana n-dimensional:

$$P = (p_1, p_2, p_3, \dots, p_n)$$

$$Q = (q_1, q_2, q_3, \dots, q_n)$$

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2 + \dots + (p_n - q_n)^2}$$