

18

Pegando os dados e marcações de teste e de treino

Anteriormente, o nosso algoritmo de classificação pegava os dados e marcações de teste da seguinte maneira:

```
treino_dados = X[:90]
treino_marcacoes = Y[:90]

teste_dados = X[-9:]
teste_marcacoes = Y[-9:]
```

Observe que, nesse caso, haviam apenas 99 dados, porém, agora estamos lidando com 1000 dados! Com certeza essa solução não funcionaria para o nosso cenário atual. Para resolvemos isso, podemos calcular a porcentagem desejada pela quantidade total de elementos. Então primeiro vamos criar a variável `porcentagem_treino` que representará a porcentagem a ser calculada, nesse caso, precisamos de 90%:

```
porcentagem_treino = 0.9
```

Agora, para pegarmos a quantidade total de elementos, basta pegarmos o tamanho de `Y` utilizando a função `len()`:

```
len(Y)
```

Para calcularmos a quantidade de treino, multiplicaremos a variável `porcentagem_treino` pelo `len(Y)`, então devolvemos para uma variável chamada `tamanho_de_treino`. Porém, além da quantidade de treino, precisamos também da quantidade de teste, para isso, basta pegar a `len(Y)` e subtrair pela variável `tamanho_de_treino` e então, atribuímos a uma variável chamada `tamanho_de_teste`. Note que dessa forma, independentemente de qual seja o valor da variável `porcentagem_treino`, sempre pegaremos o restante da porcentagem que precisamos, ou seja, se for 90% pegaremos 10%, se for 80% pegaremos 20% e assim sucessivamente.

Com as quantidades em mãos, basta pegar os valores dos arrays `X` e `Y`, mas, ao invés de `X[:90]` e `X[-9:]`, utilize as variáveis `tamanho_de_treino` para as variáveis de treino e a `tamanho_de_teste` para as de teste.

Cole o código abaixo:

Responda

INserir Código		Formatação

