

FA084 - Pre Aula 05

Use o script `pre_aula05_base.R`.

Não serão aceitos envios que não usem o script.

Entregue um arquivo `pre_aula_05_RA_primeironome.zip` contendo:

- Um arquivo `pre_aula_05_RA_primeironome.pdf` com a resposta
- Um arquivo `pre_aula_05_RA_primeironome.R` (originado do script `pre_aula05_base.R`)

O objetivo dessa pré aula é fixar a noção do split em K folds (vamos usar 5 folds). *Leia a leitura recomendada sobre cross validation*. Para isso, temos a leitura recomendada e o script para a parte prática. Essa noção deve estar bem estabelecida para aula 05 pois faremos o ajuste de parâmetros junto com o loop de cross validation. Se entenderem o conceito e implementarem o loop deste exercício, a aula será bem tranquila, pois só teremos mais linhas e colunas no loop.

Exercício 01

No script base a função `create_fold_index` tem dois argumentos `nrows` e `folds`. O primeiro é o número de linhas do `df` de entrada e o segundo é o número total de folds. Essa função já está completa. A saída dela é um vetor com o mesmo tamanho do numero de linhas do `df` que contém a sequência de 1 a 5 repetida. Por exemplo, se o `df` tivesse 12 linhas, o conteúdo da saída seria o vetor: `1 2 3 4 5 1 2 3 4 5 1 2`. *Leia a leitura recomendada sobre cross validation*. Nosso conjunto de dados tem 1309 linhas. Vale checar a consistência da saída.

Ao criar a variável `fold_index` irá conter a saída da função e será usada para indexar os folds de treino e teste em cada split do conjunto de dados. Se temos 5 folds, são 5 splits de treino e teste, 5 acurácias, 5 tpr e 5 fpr calculados.

Para desenvolver o exercício:

1. Implemente as funcoes `tpr` e `fpr`. Use os conceitos da aula 04.
2. Siga as instruções do script preenchendo o código onde necessário.

Ao fim, escreva um parágrafo (~10 linhas, flexível, não absoluto) que explique com suas palavras como o uso de cross validation impacta decisões sobre a qualidade do modelo. *Leia a leitura recomendada sobre cross validation*.

Recomendações:

Leia a leitura recomendada sobre Cross Validation

Tem 2 páginas e com figura.