

ex2_aprendizado

August 27, 2018

1 notação

Chamando a massa de dados antes da primeira seta de **estado 1**, a divisão entre treino e teste de **estado 2**, e a divisão final da validação cruzada de 5 dobras de **estado 3**

2 estado 3: validação cruzada com 5 dobras:

Cada um dos modelos (h_1 e h_2) será treinado na porção de treino (chamada de *treino* no enunciado, em Python no framework scikit-learn, esta é a etapa de fit do algoritmo) e avaliado na porção de validação (chamada de *cv* no enunciado, em Python no framework scikit-learn, esta é a etapa de predict do algoritmo) utilizando a função de custo J , em que y_i percentem ao target do conjunto de validação e \hat{h} é a saída do predict do algoritmo tendo como entrada as features do conjunto de validação

Repetimos isto para cada uma das 5 dobras para cada um dos dois modelos. Nesta etapa poderão ser encontrados os melhores hiperparâmetros de cada um dos modelos, sendo aqueles hiperparâmetros que retornarem o menor valor da função de custo

3 estado 2: “teste”

Nesta etapa, ajustamos cada um dos modelos com os dados de treino e fazemos uma previsão utilizando os dados de testes, calculando o valor da função de custo J , o melhor modelo será aquele que apresentar o menor valor de J .

Cuidar nesta etapa para não fazer um ajuste nos hiperparâmetros com base no valor J do teste, pois isto caracterizaria um sobreajustamento (overfitting) do modelo com base nos dados de treino e o modelo perderia sua capacidade de generalização

4 estado 1:

O modelo “vencedor” (que apresentar o menor J na etapa anterior) será ajustado em toda a massa de dados (chamada no enunciado de *dados*). Este modelo poderá ser salvo e utilizado para fazer previsão em novos dados não rotulados (normalmente chamados de dados de teste pela comunidade de Machine Learning ou dados em produção em Ciências da Computação)