



### Lista de Exercícios 3

1. No contexto de teste de classificadores, o que significa dizer que um estimador é enviesado? Por que o *holdout* é um estimador enviesado?
2. O *holdout* repetido é o processo no qual vários ensaios *holdout* são realizados de maneira independente uns dos outros. Esse processo possui menor viés do que um único ensaio *holdout*? Por quê?
3. Qual é o problema de se utilizar um estimador enviesado? Exemplifique uma situação na qual o uso de um estimador enviesado pode levar a prejuízo comercial ou acadêmico.
4. Elienora, uma aluna do curso de AMMD de uma turma bastante antiga, certa vez estabeleceu o seguinte procedimento de avaliação: ela construiu um conjunto com 10 mil exemplos de treinamento e partiu esse conjunto em 10 subconjuntos  $p_1, p_2, \dots, p_{10}$ . Em seguida, Elienora realizou 10 ensaios. A cada  $i$ -ésimo ensaio, ela separou o subconjunto  $p_i$  para testes e utilizou os conjuntos restantes para treinar um modelo, obtendo uma medida de erro  $e_i$ . Ao final, ela calculou a média dos erros  $e_i$  e declarou essa média como o erro provável do seu classificador.

Essa descrição é de um procedimento adequado de avaliação de classificador? Sob quais condições Elienora pode dizer que o erro do classificado será aproximadamente o valor médio encontrado?

5. Aribarbeno resolveu adaptar o método de Elienora para seus próprios propósitos. Ele fez o mesmo procedimento que Elienora utilizando o classificador SVM com kernel polinomial de grau 1. Ele chamou o erro médio encontrado de  $m_1$ . Em seguida, Aribarbeno repetiu o mesmo procedimento sobre os mesmos dados, mas com o kernel polinomial de grau 2. Ele chamou o erro médio encontrado de  $m_2$ . Ele também fez a mesma coisa para os graus 3, 4 e 5, chamando esses erros de  $m_3, m_4$  e  $m_5$ .

Aribarbeno então analisou os valores de erro médio encontrados e decidiu que  $m_3$  era o menor. Ele chegou à conclusão de que o problema possui uma complexidade razoavelmente baixa e que poderia realizar classificação com SVM polinomial de grau 3, prevendo que obterá um erro próximo de  $m_3$ .

Meses depois, Ariebarbo percebeu que seu classificador tinha um desempenho muito diferente do que havia previsto. Ficou evidente que o processo de avaliação promovido por Ariebarbo era enviesado. Quais equívocos Ariebarbo cometeu? Como ele deveria ter procedido para obter um estimador não enviesado?

6. Descreva uma situação na qual o *bootstrap* pode ser adequado quando o *10-fold cross validation* é contra-indicado.



7. Qual método de reamostragem você utilizaria para avaliar um classificador se o seu conjunto de exemplos possui apenas 100 amostras. É possível dar uma resposta para essa pergunta sem mais informações? Justifique.
  
8. Qual método de reamostragem você utilizaria para avaliar um classificador se o seu conjunto de exemplos possui um milhão de amostras. É possível dar uma resposta para essa pergunta sem mais informações? Justifique.