

Aprendizado de Máquina

Primeira Prova 2020

Nome:

1. Jon Snow está preocupado com o número de companheiros da Muralha que desertaram e fugiram para o Sul. Então ele decidiu treinar uma árvore de decisão utilizando alguns casos passados, de forma a prever deserções futuras. A tabela abaixo mostra os casos passados que Jon Snow utilizou para criar sua árvore de decisão.

Desertou?	Quantos anos na Muralha?	Já foi promovido?	Qual inclinação emocional?	De onde veio em Westeros?
0	7	não	corajoso	Campina
0	7	sim	medroso	Casterly Rock
0	8	não	medroso	Winterfell
0	6	sim	corajoso	Winterfell
0	10	sim	corajoso	Kings Landing
0	9	sim	guerreiro	Winterfell
0	6	sim	guerreiro	Terras Fluviais
1	10	não	corajoso	Terras Fluviais
1	8	não	guerreiro	Casterly Rock
1	9	não	medroso	Winterfell
1	9	não	medroso	Winterfell
1	7	não	guerreiro	Campina

- Qual é a entropia inicial de “Desertou?”
- Qual atributo seria escolhido primeiro pelo algoritmo de árvore de decisão? Por que?
- Qual o ganho de informação do atributo escolhido em (b)?
- Desenhe a árvore de decisão obtida. A árvore deve conter pelo menos três cortes ou atributos.

2. Para as próximas questões, responda com uma ou duas sentenças, ou com uma figura quando achar necessário. Para questões Verdadeiro/Falso, forneça uma justificativa caso a resposta seja Verdadeiro. Se a resposta for Falso, forneça um contra-exemplo.

V/F - Um classificador treinado em menos dados de treino é menos propenso a sobre-ajuste.

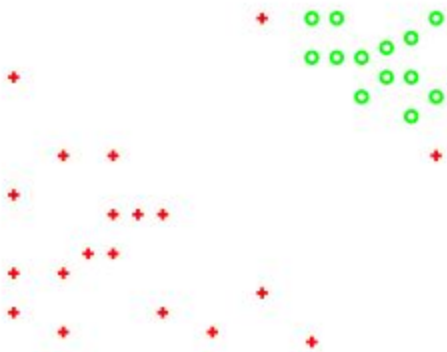
V/F - O erro de um classificador quando medido no dado de treinamento fornece uma estimativa pessimista do erro esperado.

V/F - Dados m exemplos de treinamento, o erro empírico converge ao erro esperado à medida que $m \rightarrow \infty$.

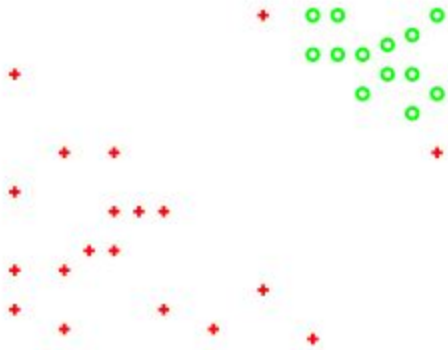
Responda: Imagine, durante a etapa de treinamento, um ponto que é corretamente classificado e fica distante da fronteira de decisão. O separador SVM seria afetado por esse ponto? E um separador Perceptron, seria? Por quê?

Responda: Por que o “true do Kernel” nos permite encontrar o separador SVM sem aumento significativo no tempo de treinamento?

3. Agora nosso objetivo é classificar corretamente os pontos de dados de teste, dado um conjunto de dados de treinamento. Esteja avisado, no entanto, que os dados de treinamento vêm de sensores sujeito a erros, portanto, você deve evitar confiar demais em qualquer ponto específico. Para este problema, suponha que estamos treinando um SVM com um kernel quadrático, ou seja, a função de kernel é um kernel polinomial de grau 2. Você recebe o conjunto de dados apresentado na Figura abaixo. A penalidade de regularização C determinará a localização do hiperplano de separação. Responda às seguintes questões qualitativamente. Dê uma resposta juntamente com a justificativa apropriada, e quando necessário desenhe sua solução na parte apropriada da Figura.
- a. Como ficaria o separador se valores altos de C fossem usados (i.e., $C \rightarrow \infty$)? Justifique.



- b. Caso voce usasse um valor de $C \approx 0$, como ficaria o separador SVM obtido? Justifique.



- c. Qual dos dois separadores/classificadores você acha que se sairia melhor em tempo de teste/classificação? Por quê?

- d. Desenhe um ponto novo o qual a inclusão mudaria o separador obtido com valores altos para C . Justifique.

