

2COP507 - Tópicos em Reconhecimento de Padrão

Prof. Sylvio Barbon Jr

Abril 2020

Atividade III - *Regressão com Máquina de Vetor de Suporte*

Teórica:

Entre os grandes problemas de reconhecimento de padrões, temos aqueles de regressão de valores. Utilizando o Aprendizado de Máquinas é possível modelar um regressor capaz de obter baixos erros preditivos. Um destes algoritmos é a Máquina de Vetor de Suporte (Support Vector Machines - SVM). Considerando a temática SVM para a indução de modelos regressores, responda as seguintes questões:

1. Qual a diferença entre predição, regressão e classificação?
2. Qual o que é o *kernel* de uma SVM?
3. Quais os tipos de *kernels* mais tradicionais e o que eles implicam nos hiperparâmetros da SVM?
4. Qual a função do hiperparâmetro custo (*cost*) da SVM?
5. Apresente duas desvantagens do uso da SVM em comparação com dois algoritmos de machine learning escolhidos por você.

Prática:

O exercício prático deste tema é o ajuste de uma função usando uma SVM para modelar a base de dados **boston** (*load_boston* - python). Para realizar este exercício, sugiro o uso da linguagem Python, com o algoritmo disponível no scikit-learn. A base de dados deve ser dividida usando uma estratégia de estratificação, use a **hold-out** com "70%" escolhidos aleatoriamente para treinar e "30%" para testar. Utilize sempre os mesmos conjuntos para as mais variadas configurações de SVM. Apresente um relatório contendo as seguintes discussões:

1. Compare o uso de 4 diferentes kernels e discuta os resultados;

2. Modifique o *cost* e discuta qual foi a influência para a base atual;
3. Encontre um kernel para modelar uma função polinomial e discuta o melhor grau encontrado.
4. Qual a métrica de erro que você escolheu para comparar os resultados? Por que?

Referência:

- [1] Bishop, Christopher M. Pattern recognition and machine learning. springer, 2006.
- [2] Witten, Ian H., and Eibe Frank. "Data mining: practical machine learning tools and techniques with Java implementations." *Acm Sigmod Record* 31.1 (2002): 76-77.