

Laboratório 3

CMC-13 Introdução à Ciência de dados

Aprendizado de Máquina Probabilístico

Aluno: Vinícius José de Menezes Pereira

Modelo Probabilístico

O modelo escolhido para o problema foi o Naives Bayes, ou seja, o modelo ingênuo, que não pressupõe uma topologia complexa de causalidade, ao contrário: considera as variáveis de atributo inicialmente independentes entre si e sem relação causal. Na topologia da rede de Bayes, portanto, temos apenas um grafo em que a variável alvo, nesse caso o voto do eleitor, aponta para as demais variáveis de atributo. Sabe-se que, porém, na realidade, essas variáveis não são tão independentes assim, pois os partidos muitas vezes sugerem uma série de temas aos eleitores, entrando todos esses temas numa caixa. Assim, quem é contra a imigração provavelmente é a favor do míssil, talvez, pois essas respostas podem estar associadas a uma recomendação proposta aos eleitores do partido 'x'. No código, criou-se um modelo com com dataset inicial.

Criação de Ranking

A criação de ranking utiliza a função Variable Elimination para, pelo cálculo da probabilidade condicional de Bayes para cada amostra do dataset de teste. Criou-se uma lista de cada amostra do teste em formato de dicionário, o que serviu de evidência para inferir as probabilidades de eleição. Após isso, as linhas foram ordenadas por ordem de maior probabilidade de votar no partido republicano.

Conclusões

Considere o laboratório proposto muito divertido e proveitoso para o aprendizado modelo de Naive Bayes de probabilidade. Mesmo sendo um modelo bastante simples em comparação com outros dispostos a utilização na ciência de dados, ele consegue ser bastante eficiente e prever muito bem o comportamento de vários fenômenos, em particular o voto de eleitores nos Estados Unidos.