

Classificação com Opção de Rejeição

Savio Lopes Rabelo

Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)
Programa de Pós-Graduação em Ciência da Computação (PPGCC)
Campus Fortaleza – CE – Brasil

savio.rabelo@ppgcc.ifce.edu.br

Resumo. Esta atividade prática estuda como a opção de rejeição afeta o desempenho do classificador Bayesiano. A metodologia utilizada para a implementação é constituída por duas fases: treinamento e teste, com cada conjunto sendo composto por 80% e 20% das bases de dados, respectivamente. Foram usadas quatro bases de dados disponíveis online no repositório UCI Machine Learning. Os resultados são bastante satisfatórios, chegando em taxas de acerto em 100% em algumas bases.

1. Introdução

No classificador Bayesiano, a adição da opção de rejeição é realizada após o processo usual de treinamento, da seguinte forma: para cada padrão de treinamento, avalia-se a classe que possui maior valor a posteriori; se o valor for menor que um limiar t , o classificador rejeita o padrão (classificando-o de acordo, caso contrário). Após essa etapa, são calculadas as taxas de rejeição e erro, respectivamente. Essas taxas são calculadas com base no dados de treinamento com objetivo de minimizar a equação proposta por [Chow 1970]. Dessa forma, essas taxas são utilizadas para escolher o limiar ótimo que deverá ser utilizado no processo de teste do classificador.

2. Simulações Computacionais

No primeiro momento foi realizada a separação do conjunto de dados em dois subconjuntos: treinamento e teste. Os valores utilizados para os conjuntos equivalem a 80% do conjunto original para a fase de treinamento e 20% do conjunto original para a fase de teste. Logo depois, os dados foram normalizados para eliminação de redundâncias indesejadas e também foram embaralhados.

Para análise comparativas neste estudo, foram usados quatro conjuntos de dados: *Iris Flower Data Set* em sua versão binária (setosa vs outras), *Vertebral Column Data Set* em sua versão binária (paciente patológico vs paciente normal), *Dermatology Data Set* (paciente patológico vs paciente normal) e *Breast Cancer Wisconsin Data Set* (paciente com câncer benigno vs paciente com câncer maligno); todos disponíveis online no repositório UCI Machine Learning [Lichman 2013].

Para realizar os experimentos, foi utilizado um computador com a seguinte configuração: processador Intel(R) Core(TM) i7-6500U a 2.5 GHz com 8 GB de RAM e executando Windows 10. Além disso, foi utilizado a linguagem de programação MATLAB. Todos os testes foram feitos com 50 realizações em cada base.

A Figura 1 resume o resultado obtido para o classificador estudados nesta atividade com a base da Coluna Vertebral, cada ponto da Figura corresponde a um valor diferente do custo de rejeição Wr , onde $Wr = [0.04 \ 0.12 \ 0.24 \ 0.36 \ 0.48]$. Para as demais bases a rejeição foi igual à zero.

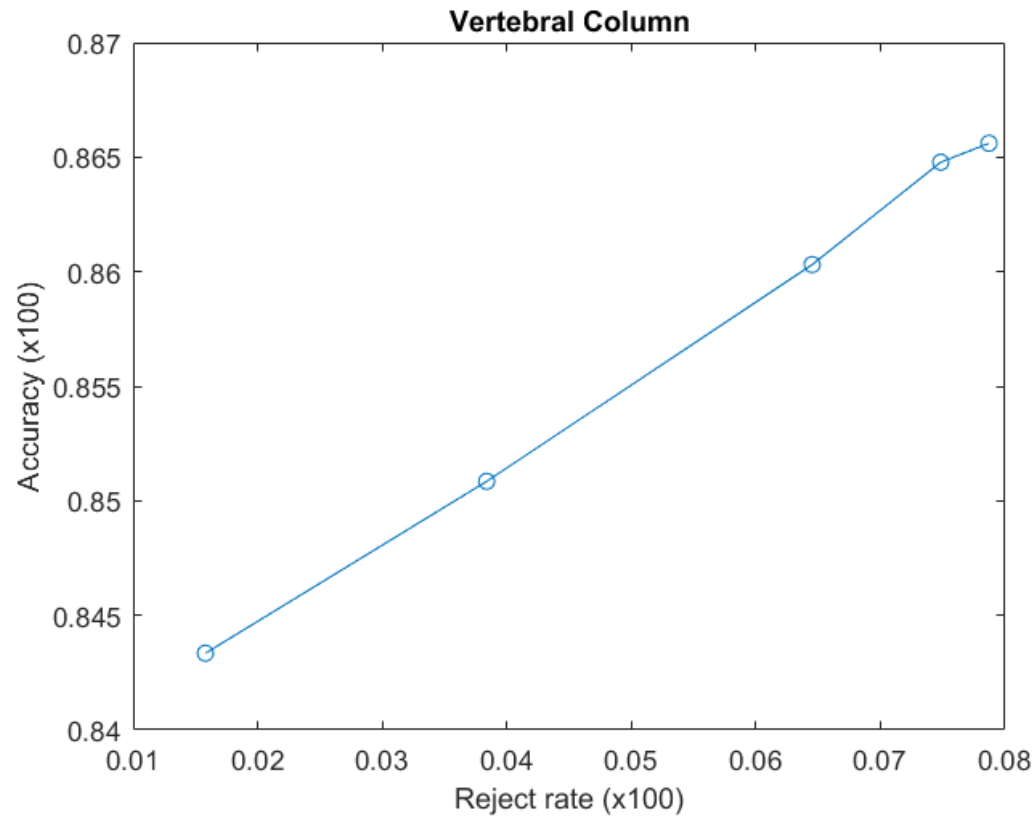


Figura 1. Curva A-R obtida pelo classificador Bayesiano

A Tabela 1 mostra os resultados do classificador Naive Bayes em todas as bases de dados.

Métricas (%)	Bases de Dados					
	Íris	Coluna (3C)	Coluna (2C)	Dermatologia	Câncer	Artificial
Acurácia	97,87	85,00	100,00	92,32	100,00	100,00
Taxa Mínima	93,33	75,81	100,00	85,14	100,00	100,00
Taxa Máxima	100,00	98,81	100,00	98,65	100,00	100,00
Desvio Padrão	02,31	05,05	00,00	02,97	00,00	00,00
Sensibilidade	97,71	80,03	100,00	91,95	100,00	100,00
Especificidade	98,91	92,56	100,00	98,58	100,00	100,00
Precisão	97,98	81,41	100,00	89,26	100,00	100,00
Tempo (s)	00,52	00,63	00,44	02,07	00,78	00,26

Tabela 1. Resultados do classificador Naive Beyes.

A Tabela 2 mostra os resultados do classificador Naive Bayes com opção de rejeição em todas as bases de dados. Lembrando que todas as bases estão em versão binária.

Métricas (%)	Bases de Dados				
	Íris	Coluna (2C)	Dermatologia	Câncer	Artificial
Acurácia	100,00	85,66	98,86	94,84	100,00
Desvio Padrão	00,00	0,68	00,00	00,00	00,00
Média de rejeição	00,00	5,10	00,00	00,11	00,00
Desvio padrão da rejeição	00,00	2,66	00,00	00,00	00,00
Tempo (s)	14,19	30,73	80,96	81,08	7,89

Tabela 2. Resultados do classificador Naive Bayes com opção de rejeição.

3. Resultados

A classificação com opção de rejeição apresenta-se muito atrativa. Em ambientes onde errar pode ser crucial, a incorporação da opção de rejeição é bastante válida pois, como apresentado pelos experimentos, pode elevar as taxas de acerto de um sistema de tomada de decisão; tanto classificando corretamente, quanto delegando essa decisão a um outro sistema ou a um especialista quando necessário.

Referências

- Chow, C. (1970). On optimum recognition error and reject tradeoff. *IEEE Transactions on information theory*, 16(1):41–46.
- Lichman, M. (2013). UCI machine learning repository. <http://archive.ics.uci.edu/ml>. Acesso em março de 2019.