

Tarefa 3 - Classificador Paramétrico

Experimento	Normalização	Seleção de Características	Projeção
1	Sim	Sim	Não
2	Não	Sim	Não
3	Sim	Não	Não
4	Não	Não	Não
5	Melhor(Sim)	Melhor(Não)	Somente a projeção (sem redução)
6	Melhor(Sim)	Melhor(Não)	Redução de Dimensionalidade

Experimento	Acurácia	95% CI	Dimensão	Parâmetros
1	90,15%	89,55%	9	54
2	58,22%	56,77%	16	89
3	91,07%	90,20%	9	54
4	56,96%	55,50%	16	89
5 (Experimento 3)	91,12%	90,26%	16	89
6 (Experimento 3)	91,23%	90,37%	6 (taxa de corte 98%)	39

Com os resultados obtidos da experimentação utilizando classificação por Misturas Gaussianas, é possível concluir que a normalização dos dados resultou em aumento de aproximadamente 54% de acurácia média na classificação dos dados. É perceptível o baixo desempenho da classificação, ao se verificar que os experimentos 2 e 4, onde a normalização dos dados não é realizada.

Visto que a normalização tem grande influência nos resultados obtidos, avaliando os desempenhos dos experimentos que utilizam o método por classe, é possível notar que todas classes obtiveram acurácias balanceadas - (*Sensitivity* + *Specificity*) / 2 - de mais de 90%.

A classe **Bombay** obteve resultados bons em todos experimentos (acurácia balanceada maior ou igual a 99,4%), o que vai de encontro com as análises e contestações realizadas nas tarefas anteriores, onde a discriminação desta classes para as demais era legível em graficos de densidade e *box-plots*. É a classe que menos apresenta amostras, mas mesmo assim seus números de *Specificity* são bons - o classificador tem baixo número de Falsos Negativos.

Classes como **Barbunya**, **Cali**, **Horoz**, **Seker** e **Sira** apresentam resultados com *Sensitivity* baixa para os experimentos sem normalização, abaixo de 60%, significando que o classificador não demonstra êxito em reconhecer estas classes em classificações binárias.

A classe **Dermason** obtém resultados razoáveis em classificações sem normalização, aproximadamente 80% de acurácia balanceada. Como destacado nas atividades anteriores, esta é a classe com mais amostras - a *Sensitivity* para esta classe apresenta porcentagem um pouco menor do que para as demais, isto acontece pois há um maior número de Falsos Positivos na classificação.

Um ponto importante para destaque é a métrica *Specificity* alta em todos experimentos, em todas as classes. Conclui-se então, que todos os classificadores identificam resultados Negativos com consistência.

O experimento 6 apresentou os melhores resultados, contando com seis dimensões e 39 parâmetros - experimento esse com menor número de dimensões e parâmetros -, mostra que a quantidade de parâmetros não sobrevalencia sobre a qualidade dos mesmos.

Considera-se que as afirmações nas análises anteriores são pertinentes para os resultados obtidos nesta tarefa. As constatações feitas ao analisar os gráficos obtidos são concretas, por exemplo para a classe **Barbunya** que devido ao seu baixo número de amostras e em gráficos de densidade para a maioria das variáveis, suas amostras estavam sendo sobrepostas por outras amostras de outras classes - mostrando então nas classificações que é uma das classes com mais dificuldade de ser identificada (*Sensitivity* baixa).