

UNIVERSIDADE FEDERAL DO AMAZONAS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DEPARTAMENTO DE APOIO À PESQUISA
PROGRAMA DE INICIAÇÃO CIENTÍFICA

Análise Discriminante Usando Densidades Preditivas Aproximadas

Bolsista: Sérgio Kardec Soares Batista, FAPEAM

Manaus - Amazonas
2016

UNIVERSIDADE FEDERAL DO AMAZONAS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DEPARTAMENTO DE APOIO À PESQUISA
PROGRAMA DE INICIAÇÃO CIENTÍFICA

RELATÓRIO FINAL
PIB-E 0098/2015-2016
Análise Discriminante Usando Densidades Preditivas Aproximadas

Bolsista: Sérgio Kardec Soares Batista, FAPEAM
Orientador: Profº Dr. Max Souza de Lima

Manaus - Amazonas
2016

Resumo

Reconhecimento e classificação de padrões são problemas importantes em uma variedade de áreas científicas, como biologia, psicologia, medicina, visão computacional e etc, mas este problema não é tão trivial quando a distribuição de probabilidade do vetor de características x é desconhecida e os componentes mesmo são correlacionados. Para o primeiro problema, na literatura, um método adequado é o BPKDA (Bayesian Predictive Kernel Discriminant Analysis) e neste projeto foi feito uma modificação nesse método para acomodar o segundo problema, ou seja, uma possível correlação entre as componentes do vetor de características x . Desta forma, apresentamos neste relatório final uma breve revisão bibliográfica do problema de reconhecimento de padrões, o método BPKDA, nossa metodologia proposta para modificação e no final uma análise sobre cada método obtido quando simulado em dados reais onde iremos observar que a modificação no método BPKDA reage bem nas simulações e em alguns casos se sai melhor que todos os métodos convencionais como veremos em uma aplicação do modelo em um problema real de classificação de fumantes ou não fumantes.

Sumário

Introdução	5
Revisão Bibliográfica	6
Métodos Utilizados	8
Resultados e Discursão	10
Conclusão	14
Agradecimentos	15
Referências Bibliográficas	16
Cronograma	17

Introdução

Reconhecimento e classificação de padrões são problemas importantes em uma variedade de áreas científicas, como biologia, psicologia, medicina, visão computacional e etc. Um padrão é representado por um vetor característica $x = (x_{i1}^\omega, x_{i2}^\omega, \dots, x_{ip}^\omega)$ e uma análise discriminante consiste em classificar um determinado padrão x^{novo} dentre r classes $\omega_1, \dots, \omega_r$ com base nas p características x_1, \dots, x_p . Para resolver esse problema existem várias maneiras de se abordar, por exemplo: rede neurais, métodos fuzzy e no nosso caso, métodos estatísticos. Do ponto de vista estatístico, assume-se que o vetor de características possui uma função densidade de probabilidade f típica de sua classe. Um vetor x pertencente a classe ω_j é visto como uma observação aleatória gerada de X de acordo com algum modelo probabilístico f condicionada à classe. Este modelo é representado por $f_j(x) = f(x|\omega_j)$.

A partir da modelagem de X é construído um sistema de reconhecimento de padrão estatístico operando em dois modos: treinamento ou aprendizagem do sistema; e classificação ou teste do sistema baseado em alguma função dos dados, chamada de classificador. Matematicamente (Melachlan, G.J., 2004), um classificador é uma função $\Phi : X \rightarrow \Omega$ tal que $\Phi(x) = \omega_j \in \Omega = \{\omega_1, \dots, \omega_r\}$. Usando uma função discriminante ϕ_j para classe ω_j (por exemplo $f_j(x)$), o classificador particiona o espaço Y de características em regiões de decisão e cada região tem a mesma classe. Os métodos estatísticos usuais de classificação (Johnson Wichern, 1998; Anderson 2003) assumem que $f_j(x)$ é conhecida (ou parcialmente conhecida) o que pode não ocorrer na prática. Então, outra maneira é estimar $f_j(x)$ de uma forma não-paramétrica (Lima and Atuncar, 2011). Assim, Lima et al, (2013) desenvolveram o BPKDA para problemas de classificação. Esse método é baseado na distribuição preditiva de um novo vetor de características x^{novo} dado um conjunto de amostras de treinamento $x^n = x_i : i = 1, 2, \dots, n$. Por fim as densidades preditivas estimadas são usadas como função discriminante considerando que os componentes do vetor x são independentes, o que nem sempre ocorre. Desta forma, propomos neste projeto uma modificação no método BPKDA de forma a acomodar a possível correlação entre os componentes do vetor de x quando estes não são independentes e principalmente sugestões onde podemos aplicar o método BPKDA modificado, pois como veremos nos resultados, o erro do método proposto será melhor que o original quando as componentes do vetor x forem correlacionadas.

Revisão Bibliográfica

Em geral os problemas de reconhecimento de padrões são classificados em duas categorias: Reconhecimento de Padrões Supervisionado(RPS) e Reconhecimento de Não Supervisionados(RPNS). O conceito de supervisionado está relacionado com o fato de ser conhecida a origem de um determinado objeto, com relação as classes envolvidas no problema abordado. Neste projeto, foi trabalhado com o caso de reconhecimento de padrões supervisionado, pois o foco é voltado para Análise de Discriminante, nossa busca é então por uma regra que funcione como uma função discriminante para a alocação de indivíduos nas classes já existentes, e isso recebe o nome de classificador. Em teoria, é possível encontrarmos uma regra ótima, mas isso requer um conhecimento completo sobre a distribuição de probabilidade dos dados, o que na nem sempre é impossível (McLachlan, 2004). Quando isso não é possível o problema pode ser contornado utilizando o método BPKDA que consiste em estimar a distribuição usando estimativas de densidades não paramétrico com base em as amostras de treinamento e para substituir estas estimativas para dar uma regra de preditiva discriminante não paramétrico. Empregando bayesiana a estrutura dada, métodos estatísticos são combinados com densidade convencional que é livre de modelo, no sentido de que não é necessária suposição sobre um modelo estatístico subjacente para o vetor de características x .

Por exemplo, no caso do Reconhecimento de Voz, sabemos que a voz se comporta como uma onda, portanto possui velocidade, período, frequência, e essa serie de fatores relevantes que vão nos da o vetor de características como sendo $x = (\omega, \tau, f, \lambda, \theta, \alpha, ..)$, onde as coordenadas do vetor são velocidade, período, frequência, comprimento de onda e etc, portanto se tivermos um banco de dados com as amostras de voz dos individuos, podemos, estabelecer limites, para dado um x^{novo} , para decidimos a que classe ele pertence, que no caso seria, a que individuo a voz pertence.

Procedimento de Classificação via BPKDA O classificador, via BPKDA, é baseado na Máxima Densidade Preditiva construído da seguinte forma: alocamos um novo indivíduo x^{novo} em uma classe $\omega_j \in \Omega$ se a classe ω_j possui a maior Densidade Preditiva $f_j(x)$ para x^{novo} . Isto é, vamos alocar x^{novo} em ω_j , tal que:

$$\omega_j = \sup_{\omega} f_{\omega}(x^{novo}|x^n) \quad (1)$$

Para a construção de $f_{\omega}(x^{novo}|x^n)$, seja $x^n = \{x_i : i = 1, 2, \dots, n\}$ vetores de características de n indivíduos (ou objetos) proveniente de um modelo probabilístico desconhecido, e cada $x_i^{\omega} = (x_{i1}^{\omega}, \dots, x_{ip}^{\omega})$, $i = 1, \dots, n$. Supondo que cada x_i^{ω} é proveniente de uma população ω_j com distribuição de probabilidade dada por $f_{\omega}(x|H)$, $H \in \mathbf{H}$, onde \mathbf{H}_{ω} é uma matriz desconhecida. Agora, dado uma no observação x^{novo} , cuja população a qual é desconhecida, podemos definir sua densidade preditiva em termos das informações x^n obtida para cada classe ω como:

$$f_{\omega}(x^{novo}|x^n) = \int_{\mathbf{H}} f_{\omega}(x^{novo}|H) \pi_{\omega}(\mathbf{H}|x^n) d\mathbf{H} \quad (2)$$

onde:

$$\pi_{\omega}(\mathbf{H}|x^n) \propto f_{\omega}(x|H) \pi_{\omega}(\mathbf{H})$$

em que α é a proporcionalidade. Após algumas manipulações algébricas e suposições sobre π_{ω} , e considerando a independência das componentes do vetor x (lima e souza, 2013) mostraram que:

$$f_{\omega}(x^{novo}|x^n) \simeq 1/k \sum_{i=1}^k \prod_{l=1}^p t \left(x_l^{novo}|x_{il}^k, \sqrt{(\bar{\beta}_l/\bar{\alpha}_l)} \right) \quad (3)$$

Onde $\bar{\beta}_l$ e $\bar{\alpha}_l$ são soluções de um sistema não linear em meio as manipulações algébricas.

Métodos Utilizados

Neste projeto, como foi dito, o BPKDA foi modificado para acomodar uma possível correlação entre os componentes do vetor de características x no processo de classificação. A diferença básica é que no BPKDA original (Lima et al.,2013) o estimador não-paramétrico de densidades é o kernel produto da distribuição unidimensional das componentes do vetor de características. Enquanto que nós iremos utilizar o kernel conjunto para estas distribuições (Scott, 1992). Desta forma, metodologicamente nossos objetivos foram alcançados da seguinte forma:

1. Inicialmente, foi modificado o BPKDA original utilizando as densidades preditivas aproximadas via kernel conjunto o qual acomoda a correlação entre o vetor de características;
2. Essas novas densidades preditivas obtidas em 1) foram utilizadas como parte de um novo classificador
3. As novas observações foram classificadas de acordo com o novo procedimento descrito em 2).

As novas densidades preditivas e o procedimento de classificação são descritos a seguir.

Estimação da Densidade Preditiva Bayesiana por Função Núcleo Multivariada Normal

Como foi citado anteriormente, nem sempre supor que as componentes são independentes, é adequado, sendo assim vamos tentar generalizar, considerando a correlação, e para obtenção de $f_{\omega}(x^{novo}|x^n)$, dada em (2), usaremos o estimador núcleo geral (Scott,1992):

$$K(x^{novo}|x_i^{(k)}, H) = \frac{1}{k} \sum_{i=1}^k |H|^{-\frac{1}{2}} K(H^{-\frac{1}{2}}(x^{novo} - x_i^{(k)}))$$

De modo que:

$$f_{\omega}(x^{novo}|\mathbf{H}) = \frac{1}{k} \sum_{i=1}^k |H|^{-\frac{1}{2}} \frac{1}{(2\pi)^{\frac{p}{2}}} \exp\left(-\frac{1}{2}(x^{novo} - x_i^k)' H^{-1}(x^{novo} - x_i^k)\right)$$

Teorema 0.0.1 *Seja $\pi_\omega(H|p, \Delta_{ij}) = \sum_{i=1}^k \omega_{ij} WI_p(H|p, \Delta_{ij})$ uma mistura de densidades Wishart Inversa (Anderson, 2003). Usando o critério da divergencia logaritima (Kullback-Leibler) temos que a melhor aproximação para mistura $\pi_\omega(H|x^{(n)})$ é*

$$WI_p(\mathbf{H}|d_j^*), \text{ com } \omega_{ij} = \frac{|\Delta_{ij}|^{-\frac{p}{2}}}{\sum_{i=1}^k |\Delta_{ij}|^{-\frac{p}{2}}}$$

onde $\mathbf{B}_j = \frac{d_j^*}{p} (\sum_{i=1}^k \omega_{ij} \Delta_{ij}^{-1})^{-1}$ e d_j^* é o valor de d_j que satisfaz:

$$\log |\mathbf{B}_j| - \sum_{v=1}^k \phi\left(\frac{d_j - v + 1}{2}\right) = \sum_{i=1}^k \omega_{ij} \log |\Delta_{ij}| - \sum_{v=1}^p \phi\left(\frac{v}{2}\right)$$

e $\Delta_{ij} = (y_j^m - x_i^k)(y_j^m - x_i^k)'$ em que $x^{(n)}$ é particinado em $x^{(k)}, y^{(m)}$ onde $k + m = n$ E ainda p é o numero de características, n é o numero de amostras de treinamento, k e m são respectivamente o numero de elementos na partição $x^{(k)}$ e $y^{(m)}$ de $x^{(n)}$. Definindo $\mathbf{B}^* = \sum_{j=1}^m \mathbf{B}_j$ e $d^* = \sum_{j=1}^m d_j^* + m - mp$, e como a priori por hipótese $f_\omega(x^{novoo}|\mathbf{H})$ tem distribuição normal e pelo teorema acima identificamos como o núcleo de uma distribuição *Wishart Inversa* e portanto temos que:

$$\mathbf{H}|x^{(n)} \sim WI_p(H|d^*, \mathbf{B}^*)$$

sendo isso, temos um resultado que garante quando temos $\mathbf{X}|\Sigma \sim N_p(\mu, \Sigma)$ e $\Sigma \sim WI_p(v, \Omega)$ então $\mathbf{X} \sim T_{1,p}(v, 2, \mu, 1, \Omega)$, ou seja, quando temos duas distribuições, Normal e Wishart Inversa multivariadas a priori, a posteriori terá distribuição t-Student multivariada após algumas manipulações algébricas nos parametros do núcleo das distribuições, teremos densidade preditiva generalizada é,:

$$f_\omega(x^{novoo}|x^n) \simeq \frac{1}{k} \sum_{i=1}^k t_p(x^{novoo}|x_i, B^*, d^*) \quad (4)$$

A expressão (4) acima, diz $x^{novoo}|x^n$ tem distribuição t-Student multivariada e acima de tudo representa uma generalização da densidade preditiva dada em (3) para o caso que os componentes do vetor de características x são correlacionados. Agora, para alocarmos um novo indivíduo $x^{(novoo)}$ em uma classe $\omega_j \in \Omega$ utilizamos a máxima densidade preditiva. Alguns importantes manipulações algebricas foram omitidas para chegarmos nesse resultado, porém poderam ser melhor observadas se o leitor acompanhar as referências.

Resultados e Discursão

Assim como qualquer método ou algoritmo, temos algumas condições para o "bom funcionamento" dos resultados obtidos, obviamente a primeira condição é que $n < \infty$, que já está subtendido, e como já reptimos exaustivamente, o método BPKDA é um classificador, digamos, "recomendado" quando as componentes do vetor x são independentes, e o classificador obtido em (4) permite componetes do vetor x sejam dependentes, o que veremos agora é que cada formula é complementar da outra para um bom funcionamento num banco de dados qualquer, em outras palavras, sabemos que o BPKDA1 erra mais que outros métodos quando o vetor x tem coorelação nas componetes, e acerta mais quando as componetes do vetor x não possui coorelação alguma, o que acontece exatamente o contrário com o classificador obtido em (4), e o chamaremos de GBPKDA, ou seja, acerta mais quando o vetor x tem coorelação nas componetes, e erra quando as componetes do vetor x não possui coorelação. Veremos agora 6 situações e um exemplo de classificação para predição de grupos de fumantes e não fumantes proposita no artigo "análise discriminate via distribuições preditivas aproximadas por estimadores por função núcleo (Souza, D.S,2013)" e em seguida, ao aplicar as modelos propostos, compararemos os erros de cada modelos, para efeito de teste, também foi aplicado em alguns modelos como BPKDA2, Análise Discriminante Linear (ADL), Análise Discriminante Quadrática (ADQ) e Naive Bayes Normal (NNBDA), que são convencionais nessa linha de problemas que não citados no trabalho, mas são usados para os mesmos propósitos.

Os métodos foram avaliados com tamanhos de amostra 100 e 300 em 1000 réplicas independentes do experimento, onde em cada réplica é obtida a taxa de erro de classificação e ao final do experimento é tomado a média dessas taxas como estimativa da taxa de erro de cada classificador. Além disso, os métodos GBPKDA, BPKDA 1 e BPKDA 2 têm dentro de seus procedimentos internos uma Validação Cruzada "Q-fold", dessa forma nesse trabalho vamos considerar 5, 7 e 9 partições internas dos dados.

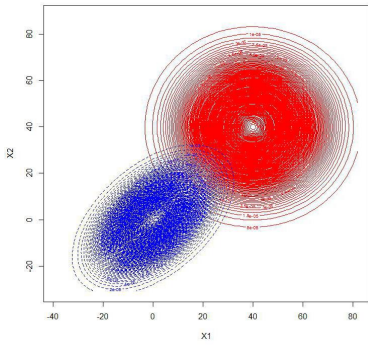


Figura 1: Estrutura 1

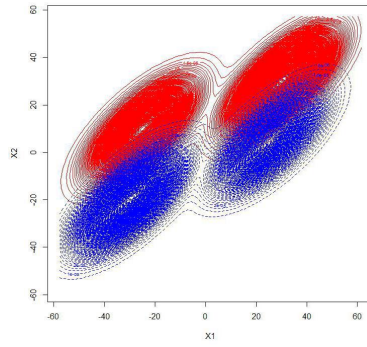


Figura 2: Estrutura 2

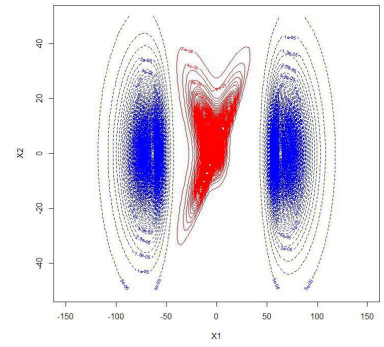


Figura 3: Estrutura 3

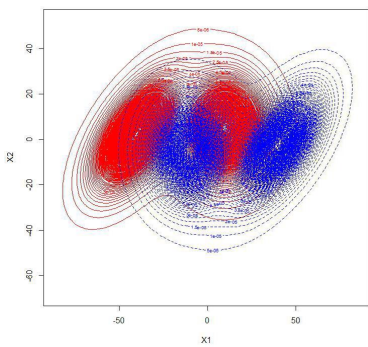


Figura 4: Estrutura 4

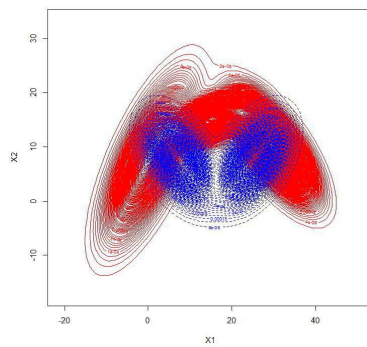


Figura 5: Estrutura 5

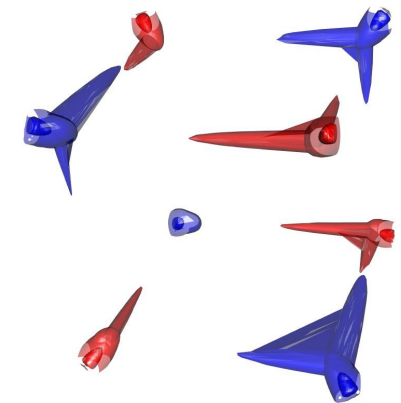


Figura 6: Estrutura 6

A seguir teremos uma tabela de erros associados a cada classificador quando testos nas estruturas acima.

Estrutura	Método	Q	N				Estrutura	Método	Q	N			
			100		300					100		300	
			Média	Sd	Média	Sd				Média	Sd	Média	Sd
1	GBPKDA	5	0.0148	0.0040	0.0139	0.0041	4	GBPKDA	5	0.2421	0.0417	0.2605	0.0302
		7	0.0158	0.0043	0.0137	0.0041			7	0.2378	0.0408	0.2391	0.0300
		9	0.0158	0.0042	0.0143	0.0041			9	0.2360	0.0406	0.2308	0.0292
	BPKDA 1	5	0.0194	0.0064	0.0180	0.0052		BPKDA 1	5	0.1880	0.0211	0.1747	0.0164
		7	0.0194	0.0061	0.0180	0.0053			7	0.1871	0.0203	0.1750	0.0160
		9	0.0194	0.0060	0.0180	0.0051			9	0.1867	0.0198	0.1762	0.0162
	BPKDA 2	5	0.0187	0.0065	0.0180	0.0054		BPKDA 2	5	0.1878	0.0222	0.1753	0.0154
		7	0.0185	0.0062	0.0178	0.0054			7	0.1872	0.0219	0.1769	0.0151
		9	0.0185	0.0061	0.0177	0.0053			9	0.1862	0.0216	0.1771	0.015
	ADL		0.0149	0.0041	0.0146	0.0039		ADL		0.2830	0.0263	0.2823	0.0216
ADQ		0.0141	0.0040	0.0128	0.0037	ADQ		0.2850	0.0272	0.2822	0.0216		
NNBDA		0.0152	0.0045	0.0143	0.0040	NNBDA		0.3047	0.0368	0.3028	0.0280		
2	GBPKDA	5	0.1003	0.0266	0.1178	0.0228	5	GBPKDA	5	0.2867	0.0488	0.3144	0.0359
		7	0.0964	0.0240	0.1082	0.0204			7	0.2939	0.0445	0.2749	0.0371
		9	0.0940	0.0226	0.1037	0.0187			9	0.2898	0.0438	0.2631	0.0332
	BPKDA 1	5	0.0845	0.0153	0.0757	0.0096		BPKDA 1	5	0.2055	0.0191	0.1919	0.0154
		7	0.0850	0.0151	0.0763	0.0098			7	0.2060	0.0189	0.1931	0.0154
		9	0.0854	0.0151	0.0767	0.0098			9	0.2066	0.0188	0.1935	0.0153
	BPKDA 2	5	0.0834	0.0149	0.0760	0.0102		BPKDA 2	5	0.2061	0.0212	0.1908	0.0163
		7	0.0839	0.0149	0.0766	0.0100			7	0.2065	0.0210	0.1917	0.0164
		9	0.0843	0.0147	0.0766	0.0099			9	0.2070	0.0211	0.1926	0.0164
	ADL		0.0658	0.0090	0.0641	0.0077		ADL		0.3703	0.0480	0.3559	0.0304
ADQ		0.0675	0.0088	0.0656	0.0079	ADQ		0.2647	0.0361	0.2423	0.0269		
NNBDA		0.2084	0.0193	0.2084	0.0145	NNBDA		0.2532	0.0308	0.2357	0.0221		
3	GBPKDA	5	0.0500	0.056	0.0299	0.0334	6	GBPKDA	5	0.1352	0.0703	0.2358	0.1586
		7	0.0539	0.0433	0.0261	0.0238			7	0.1247	0.0623	0.1051	0.0988
		9	0.0544	0.0430	0.0284	0.0173			9	0.1202	0.0583	0.0590	0.0505
	BPKDA 1	5	0.0431	0.0109	0.0328	0.0079		BPKDA 1	5	0.0450	0.0196	0.0273	0.0106
		7	0.0438	0.0104	0.0335	0.0076			7	0.0485	0.0193	0.0282	0.0105
		9	0.0444	0.0101	0.0340	0.0076			9	0.0510	0.0199	0.0287	0.0098
	BPKDA 2	5	0.0177	0.0068	0.0147	0.0052		BPKDA 2	5	0.0266	0.0134	0.0225	0.0117
		7	0.0174	0.0065	0.0146	0.0051			7	0.0276	0.0139	0.0221	0.0103
		9	0.0175	0.0060	0.0145	0.0049			9	0.0283	0.0140	0.0227	0.0105
	ADL		0.3840	0.0590	0.3883	0.0578		LDA		0.4993	0.0298	0.4992	0.0216
ADQ		0.0758	0.0726	0.0672	0.0643	QDA		0.4574	0.0762	0.4793	0.0675		
NNBDA		0.0781	0.0883	0.0698	0.0824	NNBDA		0.4991	0.0305	0.4992	0.0210		

Figura 7: Tabela 1: Média e desvio padrão das estimativa da taxa de erro de classificação.

Agora veremos outro exemplo onde todos os classificadores são mal sucedidos em um problema real na cidade de Honolulu - Havaí em 1969. Dos 7.683 indivíduos da população foram selecionados 100 e pesquisados a glicemia(mg/dL), colesterol sérico(mg/dL) e pressão sanguínea sistólica(mmHg),ou seja, um problema de 3 dimensões, e 37 desses indivíduos foram classificados como fumantes e 63 como não fumantes. Pela Figura , podemos observar que há sobreposição entre as classes, uma correlação positiva considerável entre as variáveis, e também há presença de observações discrepantes. Tal estrutura de sobreposição pode implicar em ambiguidade na classificação que causa grandes problemas para os métodos de classificação, principalmente para aqueles que se baseiam na Aplicação em dados reais distância entre as observações, que é o caso dos procedimento por função núcleo, mas também podem implicar em dificuldade para os métodos clássicos, pois os centroides das classes estão muito próximos como veremos a seguir

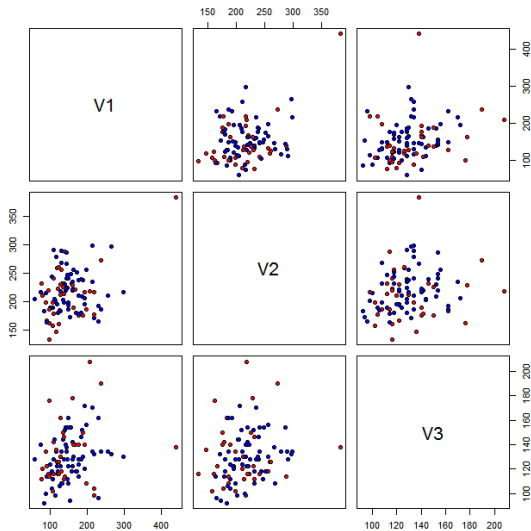


Figura 8: Por pares de variáveis

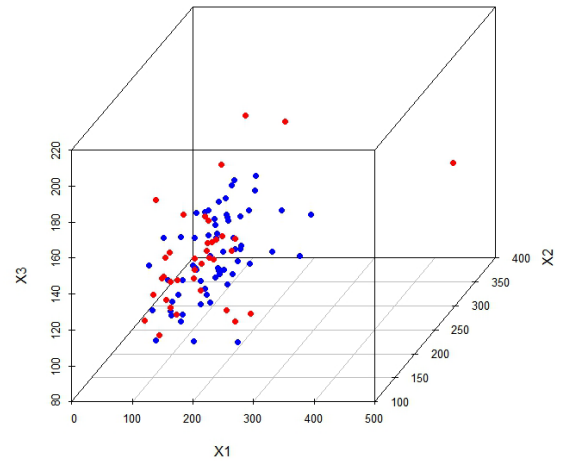


Figura 9: Em 3 dimensões

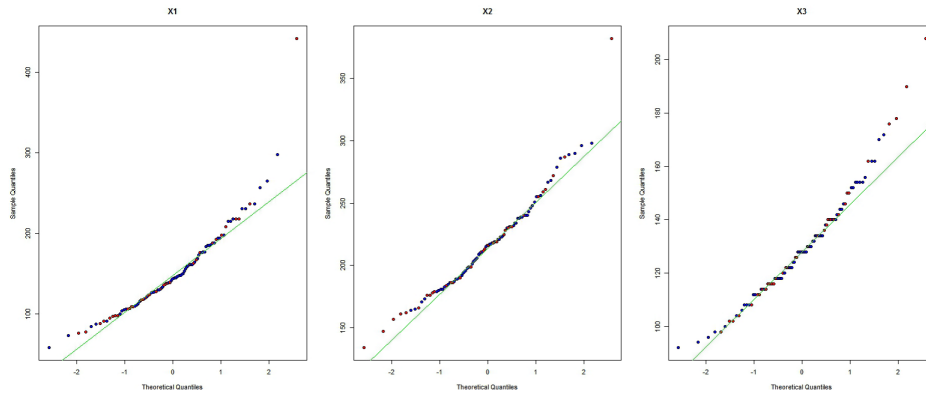


Figura 10: Gráfico Q-Q Normal das variáveis de Honolulu

Pela Figura 10 , observamos de modo independente que as variáveis provavelmente não possuem normalidade. agora novamente veremos a tabela de erros associado a cada modelo quando aplicados no problema em questão: observa-se na tabela que quando aplicado cada

Method	Q	Erros
GBPKDA	5	0.3600
	7	0.3600
	9	0.3400
BPKDA1	5	0.5300
	7	0.5300
	9	0.5100
BPKDA2	5	0.5000
	7	0.4900
	9	0.5200
ADL	NÃO SE APLICA	0.4400
ADQ	NÃO SE APLICA	0.4200
NNBDA	NÃO SE APLICA	0.4800

Figura 11: tabela de erros

classificador usado para classificar o primeiro exemplo no exemplo em questão, vemos que nosso GPKDA tem a menor taxa de erros para esse caso, com 34% de erro quando usado

Q=9 partição interna dos dados, que obviamente não é um resultado satisfatório, entretanto os outros classificadores conseguem ser ainda pior como por exemplo o NNBDa que teve 46% de erro.

Conclusão

Neste trabalho, apresentamos duas novas abordagens não paramétricas de Análise Discriminante baseadas em estimadores por funções núcleo adotando uma abordagem Bayesiana, a Densidade Preditiva Bayesiana por Função Núcleo Multivariada Normal (GBPKDA), a Densidade Preditiva Bayesiana por Produto de Funções Núcleo Normais (BPKDA 1). E como já foi antecipado nos resultados, observa-se em alguns testes nas estruturas, cada classificador é indicado para casos diferentes, ou seja, onde usa-se um, não se usa o outro, por razões óbvias, entretanto, de um modo geral, considerando todos os resultados obtidos nas simulações, o método que mais se destacou foi o BPKDA 2 por sua versatilidade, pois obteve a menor taxa média de erros, porém no segundo exemplo como vimos, os resultados foram ruins para todos os métodos, a menor taxa de erro foi do nosso GBPKDA com pouco mais de 33% (um absurdo quando queremos acertar na grande maioria das vezes). Uma maneira interessante de implementar algo que fosse melhor que qualquer algoritmo em específico é usar os três pacotes em um só algoritmo fazendo if/else's para cada caso rodando classificadores diferentes e escolhendo, é claro, um dos quais apresentar o menor erro em cada caso, e fica em aberto para algum novo classificador melhorar o que não conseguimos acertar como no caso do exemplo dos fumantes na cidade de Honolulu no Havaí, onde os centroides das classes estão muito próximos, e como sabemos e foi dito nesse relatório, nosso modelo particiona o espaço em regiões de decisão que tenham mesma classe, como vemos nesse último exemplo, há uma sobreposição dos limites de decisão, isso se deve ao fato de que foram fornecidos poucos dados, e com isso podendo haver interseções nas características de fumantes ou não fumantes, como por exemplo o colesterol que é alto em fumantes mas também nos não fumantes que tenha maus hábitos alimentares assim como a pressão sanguínea sistólica, e isso justifica o mau desempenho de todos os métodos e exalta o fato de mesmo com poucos dados nosso método ainda se sair melhor que os demais.

Agradecimentos

Primeiramente, agradeço ao meu orientador, Prof. Dr. Max de Souza Lima pela paciência na minha plena ignorância e por ter feito um ótimo trabalho e me orientado é claro, ao apoio financeiro concedido pela FAPEAM.

Referências Bibliográficas

- Anderson, T. W. (2003), An introduction to multivariate statistical analysis, 3rd. edn, Wiley-Interscience, New Jersey...
- McLachlan, G. J. (2004). Discriminant Analysis and Statistical Pattern Recognition . John Wiley & Sons, NY..
- Johnson, R. A., Wichern, D. W. (1998), Applied multivariate statistical analysis, 4th. edn, Prentice-Hall, New York. . Lima, M.S, Atuncar, G.S. (2011) A Bayesian method to estimate the optimal bandwidth for multivariate kernel estimator . Journal of Nonparametric Statistics, v.23(1), p. 137- 148.
- Lima, M.S, PEREIRA., J.R.G., SILVA., S.D. (2013) Bayesian predictive kernel discriminant analysis. Pattern Recognition Letters, v. 34, p. 2079-2085. Scott, D. (1992) Multivariate Density Estimation: Theory, Practice and Visualization. [S.l.]: New York: John Wiley.

Cronograma

Descrição	Ago 2015	Set	Out	Nov	Dez	Jan 2016	Fev	Mar	Abr	Mai	Jun	Jul
Revisão da Literatura sobre o BPKDA.	R	R	R	R	R							
Implementação da modificação no algoritmo BPKDA já desenvolvido em R;			R	R	R							
Avaliação do método BPKDA modificado, através de dados simulados.						R	R	R				
Avaliar o desempenho dessa nova versão do BPKDA em conjuntos de dados reais.									R	R	R	
Elaboração do Resumo e Relatório Final - Preparação da Apresentação Final para o Congresso												R

Figura 12: Gráfico R é realizado e F é à finalizar