Error PCA: A Density-Overlap Approach to Supervised Feature Extraction for Classification

Cloves Cabral Vilela Filho[1], **Anderson Pinheiro Cavalcanti**[2]
Tiago Buaque A. de Carvalho[1]

[1]Universidade Federal do Agreste de Pernambuco (UFAPE)
[2]Universidade Federal Rural de Pernambuco (UFRPE)

cloves.cabral46@gmail.com

anderson.pinheiro@ufrpe.br

tiago.buarque@ufape.br

Abstract. *Principal Component Analysis (PCA) is well-established as one of the most widely used techniques for feature extraction. This work introduces Error PCA (EPCA), a supervised approach that integrates similarity measures, specifically, probability distribution overlap coefficients, into the classical PCA framework. Experimental validation was conducted by comparing the accuracy of EPCA against traditional PCA, MCEPCA, and Supervised PCA (SPCA), evaluated across four classification algorithms using multiple benchmark datasets. The results demonstrate that the proposed method matched or outperformed the benchmark techniques in terms of predictive accuracy.*

Resumo. *A Análise de Componentes Principais (PCA — Principal Component Analysis) consolida-se como uma das técnicas mais difundidas para a extração de características. Este trabalho apresenta o EPCA (Error PCA), uma abordagem supervisionada que integra medidas de similaridade, especificamente, coeficientes de sobreposição de distribuições de probabilidade, ao PCA clássico. A validação experimental foi conduzida por meio da comparação da acurácia do EPCA com o PCA tradicional, o MCEPCA e o PCA Supervisionado (SPCA), avaliados sob quatro algoritmos de classificação em múltiplos conjuntos de dados. Os resultados demonstram que o método proposto igualou ou superou as técnicas comparadas em termos de acurácia preditiva.*

1. Introdução

A alta dimensionalidade dos dados é um obstáculo recorrente em aprendizado de máquina: muitos atributos elevam o custo computacional, aumentam o tempo de treinamento e podem degradar a acurácia dos classificadores. Em espaços de alta dimensão, as distâncias entre exemplos tendem a se uniformizar, fenômeno conhecido como maldição da dimensionalidade, que dificulta a generalização [Webb 2002]. Técnicas de redução de dimensionalidade mitigam esse efeito descartando informação redundante e preservando a estrutura relevante dos dados [Fukunaga 1990].

A Análise de Componentes Principais (PCA) [Jolliffe 2002] [chin Ivan Chang 2026] [Vogelstein 2021] é o método clássico mais empregado nessa tarefa. O PCA produz componentes não correlacionados, ordenados pela variância, mas é *não supervisionado*: ignora os rótulos, de modo que a direção de maior variância

nem sempre é a mais discriminante. Para incorporar a informação de classe surgiram variantes supervisionadas. O Supervised PCA (SPCA) [Barshan et al. 2011] busca uma transformação ortogonal cujos componentes maximizem a dependência com a variável resposta. O Minimum Classification Error PCA (MCEPCA) [de Carvalho et al. 2017] ordena os componentes por um *score* inversamente proporcional ao erro bayesiano. O MCEPCA, contudo, impõe restrições fortes: assume normalidade, duas classes, probabilidades *a priori* iguais, mesma matriz de covariância e características independentes.

Neste trabalho propomos o EPCA (*Error PCA*), um método supervisionado de seleção de características que substitui o *score* paramétrico do MCEPCA por uma medida direta de sobreposição entre as densidades de cada classe: o coeficiente de sobreposição (OVL) [Weitzman 1970, Inman and Bradley Jr. 1989]. A intuição é que, quanto menor a área de interseção entre as densidades de duas classes em um dado componente, menor o erro de classificação associado a esse componente. Ao operar diretamente sobre as densidades, o EPCA dispensa duas restrições do MCEPCA: passa a tratar problemas com mais de duas classes e não exige covariâncias iguais.

Organizamos o estudo em torno de três questões de pesquisa. **(QP1)** Em problemas binários, o EPCA mantém o desempenho do MCEPCA? **(QP2)** Em problemas multiclasse, para os quais o MCEPCA não se aplica, o EPCA seleciona componentes mais discriminantes que o PCA e o SPCA? **(QP3)** Quais restrições do MCEPCA o EPCA de fato elimina e quais permanecem? As Seções seguintes apresentam a fundamentação, o método proposto, a metodologia experimental, os resultados, as limitações e as conclusões.

2. Fundamentação

A redução de dimensionalidade supervisionada para classificação é um tema clássico em reconhecimento de padrões. A Análise de Discriminante Linear (LDA) busca projeções que maximizem a razão entre as dispersões interclasse e intraclasse, mas assume densidades gaussianas com covariância comum e produz, no máximo, $l - 1$ componentes para l classes [Fukunaga 1990], o que limita sua aplicação quando poucas projeções não capturam toda a estrutura discriminante. O PCA, por sua vez, é ótimo em variância, porém indiferente aos rótulos [Jolliffe 2002].

Entre as variantes supervisionadas do PCA, o SPCA [Barshan et al. 2011] formula a seleção de subespaço como a maximização da dependência estatística (via critério de Hilbert-Schmidt) entre os dados projetados e os rótulos, com solução em forma fechada e versão kernelizada para problemas não lineares. O MCEPCA [de Carvalho et al. 2017] adota uma perspectiva orientada ao erro: após descorrelacionar os dados com o PCA, atribui a cada componente um *score* derivado de uma aproximação do erro bayesiano sob hipóteses gaussianas, selecionando os componentes de menor erro. O EPCA situa-se nessa mesma linha orientada ao erro, mas difere ao estimar o erro *diretamente* pela área de sobreposição das densidades de cada classe, o que dispensa a forma paramétrica fechada do MCEPCA e, com isso, o limite de duas classes e a exigência de covariâncias iguais.

2.1. Análise de Componentes Principais (PCA)

O PCA [Jolliffe 2002] gera componentes não correlacionados, combinações lineares das variáveis originais, ordenados pela variância. Dada uma matriz $X_{n\times d}$ com n exemplos e d

atributos, calcula-se o vetor de médias $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$, centralizam-se os dados ($x = X - \bar{x}$) e obtém-se a matriz de covariância $\Sigma = \frac{1}{n}x^\top x$. Os autovetores $E = [e_1, \ldots, e_d]$ de Σ e seus autovalores $A = [a_1, \ldots, a_d]$ definem as novas direções; ordenando os autovetores do maior para o menor autovalor e projetando $W = xE_k$, obtém-se a representação reduzida em $k \leq d$ dimensões.

2.2. Supervised PCA (SPCA)

O SPCA [Barshan et al. 2011] procura uma transformação ortogonal que maximize a dependência entre os componentes e os rótulos. Define-se

$$Q = X\,H\,L\,H\,X^\top, \qquad H = I - n^{-1}ee^\top, \tag{1}$$

em que H é a matriz de centralização ($ee^\top$ é a matriz $n \times n$ de uns) e L é um *kernel* dos rótulos. Constrói-se $l_{c\times n}$ com $l_{ij} = 1$ se o exemplo j pertence à classe i e 0 caso contrário, e toma-se $L = l^\top l$. Os autovetores de Q, ordenados pelo maior autovalor, definem a projeção, de forma análoga ao PCA.

2.3. Minimum Classification Error PCA (MCEPCA)

O MCEPCA [de Carvalho et al. 2017] é supervisionado e busca minimizar o erro bayesiano. Para tornar a estimativa tratável, impõe que: (1) os dados são normais; (2) há apenas duas classes; (3) as probabilidades *a priori* são iguais; (4) as classes compartilham a matriz de covariância; e (5) as características são independentes. Sob essas hipóteses, aplica-se primeiro o PCA (para obter componentes descorrelacionados) e atribui-se a cada componente i um *score*

$$s_i = \frac{(\bar{w}_{1i} - \bar{w}_{2i})^2}{\lambda_i} \quad (\lambda_i \neq 0), \tag{2}$$

em que λ_i é o autovalor e $\bar{w}_{ci}$ a média do componente i na classe c. Como o *score* é inversamente proporcional ao erro, selecionam-se os componentes de maior s_i.

2.4. Densidades e coeficiente de sobreposição (OVL)

Uma função de densidade de probabilidade (FDP) $f(x)$ associa a cada valor de uma variável aleatória contínua sua densidade, com $f(x) \geq 0$ e $\int_{-\infty}^{\infty} f(x)\,dx = 1$ [Brownlee 2020, DasGupta 2011]. O coeficiente de sobreposição (OVL) mede a similaridade entre duas FDPs, variando de 0 (sem interseção) a 1 (densidades idênticas) [Inman and Bradley Jr. 1989, Linacre 1996]. Entre as medidas propostas na literatura (Matusita, Morisita, Kullback–Leibler e Weitzman), seus valores costumam ser próximos [Jose et al. 2019, Dhaker et al. 2019]; adotamos a medida de Weitzman Δ (Figura 1), definida como a área de interseção das duas densidades, com implementação disponível no módulo `statistics` do Python[1]:

$$\Delta = \int \min[f_1(x),\, f_2(x)]\; dx. \tag{3}$$

Para duas classes equiprováveis, o erro de Bayes é exatamente metade dessa área, $\varepsilon_{\text{Bayes}} = \frac{1}{2}\int \min[f_1, f_2]\,dx = \frac{1}{2}\Delta$ [Inman and Bradley Jr. 1989]. Portanto, ordenar componentes

[1] `https://github.com/python/cpython/blob/3.8/Lib/statistics.py`

por Δ crescente equivale a ordená-los por erro de Bayes crescente, o que justifica o uso do OVL como critério de seleção.

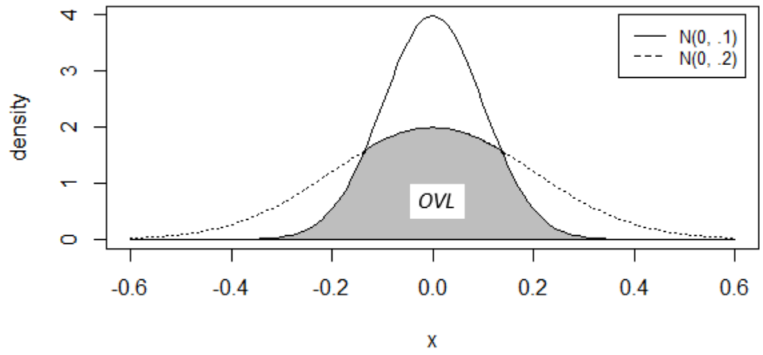


Figura 1. Sobreposição de duas distribuições normais [Dhaker et al. 2019].

2.5. Classificadores

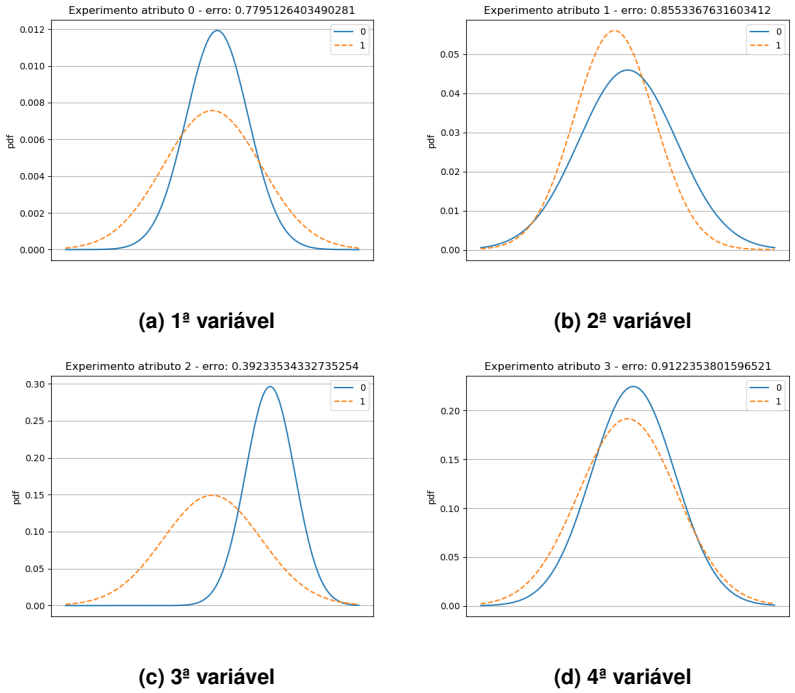
Avaliamos a qualidade dos componentes selecionados com quatro classificadores de naturezas distintas [Mitchell 1997, Webb 2002]. O k-NN classifica cada exemplo pelo voto majoritário de seus k vizinhos mais próximos, sendo sensível à geometria local do espaço de características. A Árvore de Decisão particiona recursivamente o espaço por testes em atributos, formando regras hierárquicas interpretáveis. O Discriminante Linear assume densidades gaussianas com covariância comum e separa as classes por fronteiras lineares. O Naive Bayes aplica o teorema de Bayes supondo independência condicional entre os atributos dada a classe. Essa diversidade permite verificar se os componentes selecionados pelo EPCA são úteis para classificadores com vieses indutivos diferentes, e não apenas para um deles.

3. Método Proposto: EPCA

O EPCA seleciona os componentes principais (autovetores do PCA) com menor erro estimado. Diferentemente do MCEPCA, que usa o *score* paramétrico da Equação 2, o EPCA estima o erro de cada componente diretamente pela área de sobreposição das densidades de cada classe. Sendo $f_0(X_i)$ e $f_1(X_i)$ as FDPs das classes 0 e 1 no componente i, e d o número de componentes, define-se o erro do componente i para duas classes como

$$erro_i = \mathrm{OVL}(f_0(X_i), f_1(X_i)), \quad 0 \leq i \leq d. \tag{4}$$

Como a Equação 4 não pressupõe igualdade de covariâncias, o EPCA já remove essa restrição do MCEPCA. A Figura 2 ilustra o critério na base *Banknote*: para cada componente mostramos as densidades das duas classes e o erro (área de sobreposição) correspondente. A terceira variável apresenta a menor sobreposição e, portanto, o menor erro, separando melhor as classes, sendo a primeira a ser selecionada pelo EPCA.



(a) 1ª variável **(b) 2ª variável**

(c) 3ª variável **(d) 4ª variável**

Figura 2. FDPs das classes para os atributos da base *Banknote* com o erro (área de sobreposição) de cada componente.

3.1. Extensão para mais de duas classes

Para l classes, calcula-se o erro de cada componente como a média das sobreposições de todas as $\binom{l}{2}$ combinações de pares de classes. Sendo e_{ij} o erro (Equação 4) do par j,

$$erro_i = \frac{1}{\binom{l}{2}} \sum_{j=1}^{\binom{l}{2}} e_{ij}, \quad 0 \leq i \leq d. \tag{5}$$

Essa agregação por pares é o que permite ao EPCA tratar problemas multiclasse, removendo a segunda restrição do MCEPCA.

3.2. Algoritmo

O treinamento aplica o PCA ao conjunto de treino, separa os dados por classe e, para cada componente, ajusta uma FDP normal (por média e desvio padrão de cada classe), computa o erro pela Equação 5, reordena os autovetores do menor para o maior erro e projeta os dados como podemos ver no pseudo código (Algoritmo 1). O ajuste das FDPs e o cálculo do OVL utilizam *apenas* o conjunto de treino, o que evita vazamento de informação do teste para a seleção.

Algorithm 1 EPCA – seleção de componentes por sobreposição

$X_{\text{tr}}, Y_{\text{tr}} \leftarrow$ DADOSDETREINO()
$Z \leftarrow \text{PCA}(X_{\text{tr}})$ ▷ componentes descorrelacionados
$G \leftarrow$ SEPARARPORCLASSE(Z, Y_{tr})
$erros \leftarrow$ ZEROS(d)
for $i \leftarrow 1$ **to** d **do**
    for $(a, b) \in$ PARESDECLASSES() **do**
        $f_a \leftarrow \text{FDP}(G[a].\text{componente}[i])$
        $f_b \leftarrow \text{FDP}(G[b].\text{componente}[i])$
        $erros[i] \leftarrow erros[i] + \text{OVL}(f_a, f_b)$
    end for
    $erros[i] \leftarrow erros[i]/totalCombinações$
end for
$aut \leftarrow$ ORDENARPORERROCRESCENTE(PCA.AUTOVETORES, $erros$)
$W \leftarrow$ PROJETAR(X, aut)

3.3. Complexidade computacional

Sobre o custo do PCA, o EPCA acrescenta o cálculo de $\binom{l}{2}$ coeficientes de sobreposição por componente. Para d componentes e l classes, são $d \cdot \binom{l}{2}$ avaliações de OVL. Modelando as FDPs como normais e aproximando o coeficiente de Weitzman por integração numérica em uma grade de g pontos, cada avaliação custa $O(g)$, de modo que o custo adicional do EPCA é $O(d\, l^2\, g)$. O termo quadrático em l domina o custo em problemas com muitas classes, na base *Letter*, com $l = 26$, são $\binom{26}{2} = 325$ pares por componente, o que motiva, como trabalho futuro, uma medida de sobreposição definida diretamente para múltiplas densidades.

4. Metodologia Experimental

Tabela 1. Bases de dados utilizadas (UCI).

Base	Exemplos	Atributos	Classes	Tipo
Bank	45.000	48†	2	Binária
Banknote	1.372	4	2	Binária
Occupancy	20.000	5	2	Binária
Pima	768	8	2	Binária
VColumn	310	6	2	Binária
Dermatology	366	34	6	Multiclasse
Letter	20.000	16	26	Multiclasse
Mice	1.080	83†	8	Multiclasse
Wine	178	13	3	Multiclasse
Iris	150	4	3	Multiclasse

† Após codificação *one-hot* dos atributos categóricos.

Comparamos o EPCA com o PCA clássico, o SPCA e o MCEPCA. Em problemas binários, os quatro métodos são avaliados; em problemas multiclasse, o

MCEPCA é omitido por não ser aplicável. Foram usadas dez bases do repositório UCI [Dua and Graff 2019], cinco binárias e cinco multiclasse (Tabela 1). Atributos categóricos foram convertidos em numéricos por codificação *one-hot* antes da extração, o que explica o aumento do número de atributos em *Bank* e *Mice*.

Cada configuração foi avaliada com 100 *holdouts* aleatórios (50% treino, 50% teste) Essa configuração, embora simples, garante clareza na separação entre treino e validação e tem sido discutida em trabalhos recentes como uma alternativa válida. [MCALINN K. 2026]; reportamos a média da acurácia. Usamos quatro classificadores, k-NN, Árvore de Decisão, Discriminante Linear e Naive Bayes, com as implementações e hiperparâmetros padrão do *scikit-learn* [Pedregosa et al. 2011]. As FDPs foram modeladas como normais (média e desvio padrão por classe) e o coeficiente de Weitzman foi obtido por integração numérica. O mesmo protocolo e os mesmos classificadores foram aplicados a todos os métodos, garantindo comparação controlada.

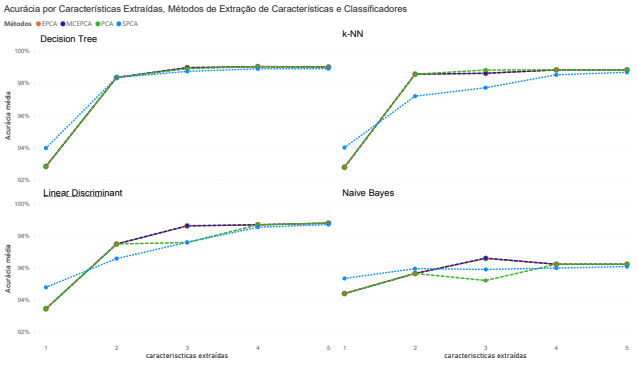
5. Resultados e Discussão

5.1. Problemas binários (QP1)

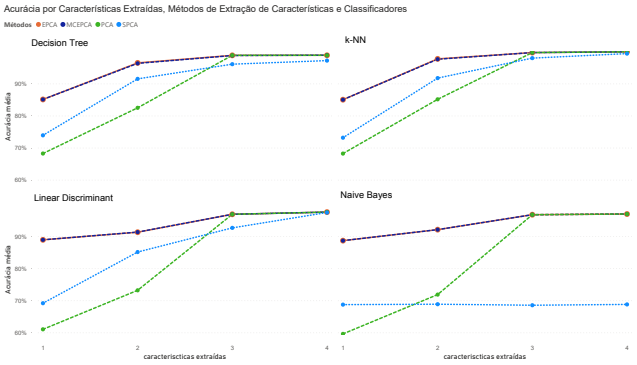
Nas cinco bases binárias, o EPCA apresentou comportamento muito próximo ao do MCEPCA (Tabela 2). Em *Bank*, na primeira componente, o EPCA atingiu 88,81% com o Discriminante Linear contra 88,9% do MCEPCA, diferença inferior a 0,1 ponto percentual. Em *Banknote* (Figura 3b), as curvas dos dois métodos coincidiram. Em *Occupancy* (Figura 3a), o SPCA descreveu melhor a primeira componente ($>$94% em todos os classificadores), mas EPCA e MCEPCA o superaram a partir da segunda ($>$97%). Em *Pima* (Figura 4), o EPCA superou 72% com Naive Bayes e Discriminante Linear, e em *VColumn* alcançou 82,32% com quatro componentes. O conjunto indica que, em problemas binários, o EPCA *equipara-se* ao MCEPCA, respondendo afirmativamente à QP1.

Tabela 2. Melhor resultado do EPCA por base binária e comparação qualitativa.

Base	Classificador	Acur. EPCA	Comparação
Bank	Discriminante Linear	88,81% ± 0.55% (1 comp.)	$\approx$ MCEPCA (88,9%)
Banknote	–	= MCEPCA	curvas coincidentes
Occupancy	todos	$>$97% ± 0.44%(2 comp.)	$>$ SPCA a partir de 2 comp.
Pima	NB / Disc. Linear	$>$72% ± 2%(1 comp.)	$\approx$ MCEPCA
VColumn	Discriminante Linear	82,32% ± 4%(4 comp.)	$\geq$ demais métodos



(a) *Occupancy*.



(b) *Banknote*.

Figura 3. Acurácia por número de componentes em duas bases binárias. EPCA e MCEPCA têm comportamento praticamente idêntico.

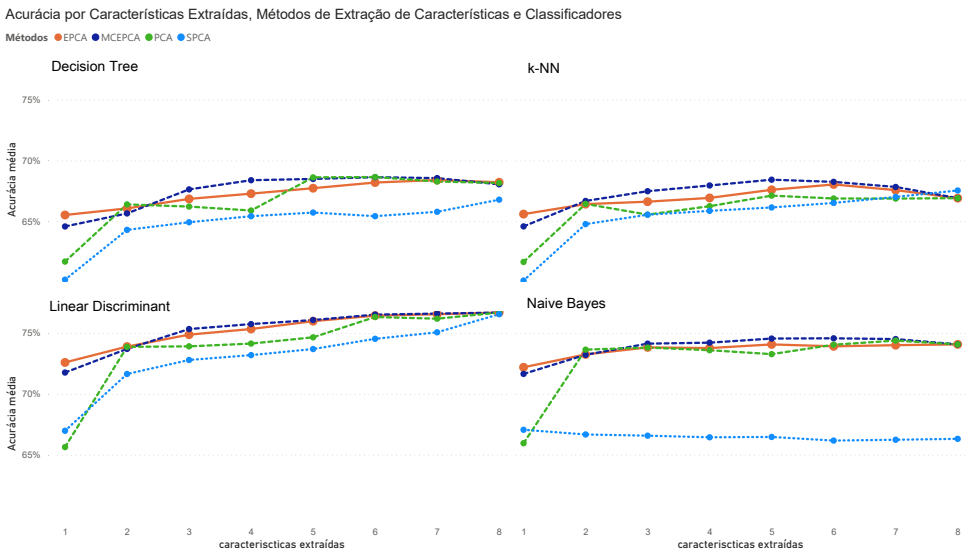


Figura 4. Acurácia por número de componentes na base *Pima*.

5.2. Problemas multiclasse (QP2)

Nas bases multiclasse, o EPCA superou o PCA e o SPCA na maioria dos casos (Tabela 3). Em *Dermatology* (Figura 5a), o EPCA ultrapassou 70% em todos os classificadores, enquanto PCA e SPCA ficaram próximos de 40% e 30%, atingindo 89,83% com k-NN e cinco componentes. Em *Letter* (Figura 6a), superou 96% com k-NN a partir de 14 componentes; em *Mice*, ultrapassou 90% com k-NN e duas componentes; em *Wine* (Figura 6b), alcançou 96,26% com Discriminante Linear e onze componentes. A exceção foi *Iris* (Figura 5b): o SPCA obteve mais de 93% com uma única componente, enquanto o EPCA empatou com o PCA em 92%. Atribuímos esse resultado à agregação por pares (Equação 5): a média das sobreposições pode mascarar uma componente que separa bem um par de classes mas mal os demais, levando a uma estimativa de erro menos precisa. Esse caso evidencia uma limitação intrínseca da estratégia de média, discutida na Seção 6. No conjunto, os resultados respondem afirmativamente à QP2 na maioria das bases, com a ressalva de *Iris*.

Tabela 3. Melhor resultado do EPCA por base multiclasse e comparação qualitativa.

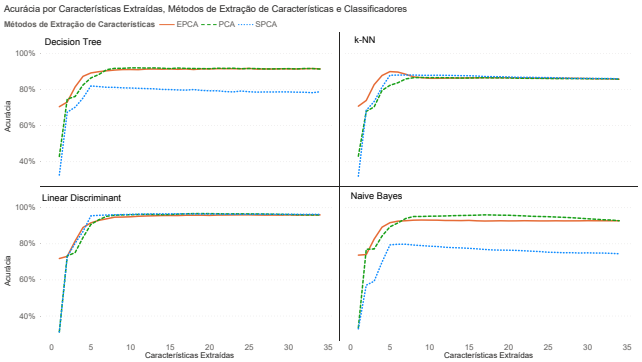
Base	Classificador	Acur. EPCA	Comparação
Dermatology	k-NN	89,83% ± 4.2%(5 comp.)	$\gg$ PCA ($\sim$40%), SPCA ($\sim$30%)
Letter	k-NN	$>$96% ± 0.22%(14 comp.)	$>$ PCA e SPCA
Mice	k-NN	$>$90% ± 3.4%(2 comp.)	$>$ PCA e SPCA
Wine	Discriminante Linear	96,26% ± 1.7%(11 comp.)	$>$ PCA e SPCA
Iris	–	92% ± 2%(1 comp.)	$=$ PCA; $<$ SPCA ($>$93%)

5.3. Restrições efetivamente removidas (QP3)

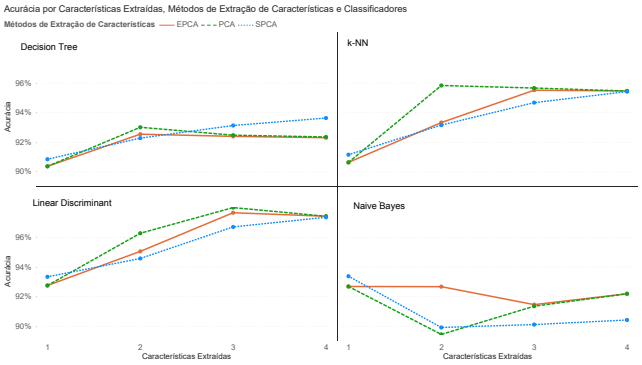
Quanto à QP3, o EPCA elimina duas das cinco restrições do MCEPCA: o limite de duas classes (via Equação 5) e a exigência de matrizes de covariância iguais (pois o OVL opera diretamente sobre as densidades). Permanecem, contudo, três hipóteses: a normalidade das densidades (assumida no ajuste das FDPs), a independência entre componentes (herdada do PCA aplicado previamente) e a igualdade das probabilidades *a priori* (implícita na média simples dos pares). Essas hipóteses remanescentes orientam as limitações e os trabalhos futuros.

5.4. Discussão geral

Os experimentos revelam um padrão coerente. Em problemas binários, EPCA e MCEPCA praticamente se confundem, o que é esperado: ambos perseguem o mesmo objetivo, minimizar o erro de Bayes e, para duas classes, o coeficiente de sobreposição é proporcional a esse erro. A vantagem do EPCA manifesta-se nos problemas multiclasse, nos quais o MCEPCA sequer se aplica e o EPCA seleciona, na maioria das bases, componentes mais discriminantes que o PCA e o SPCA. Chama atenção que poucos



(a) *Dermatology*: EPCA supera amplamente PCA e SPCA.



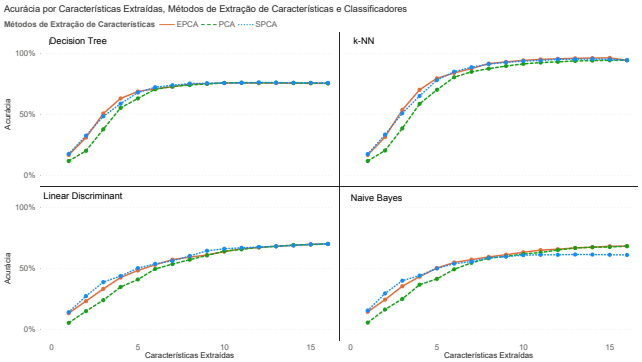
(b) *Iris*: caso desfavorável ao EPCA.

Figura 5. Acurácia por número de componentes em duas bases multiclasse.

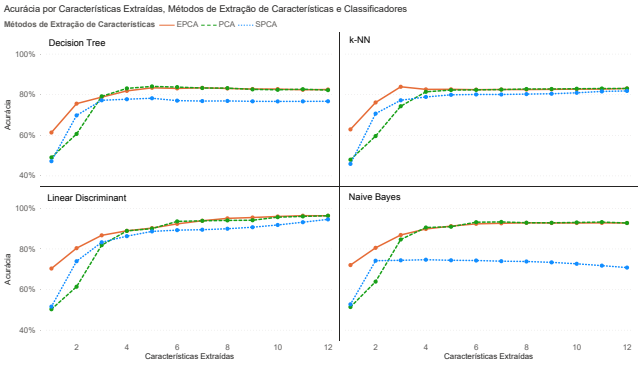
componentes bastam para atingir alta acurácia em várias bases (duas em *Mice*, cinco em *Dermatology*), indicando que o critério baseado em sobreposição é eficaz já nas primeiras posições da ordenação. O caso desfavorável de *Iris* é informativo: trata-se de uma base de baixa dimensionalidade (quatro atributos, três classes), em que a média sobre apenas três pares é grosseira e pode não refletir a componente globalmente mais discriminante. Esse comportamento reforça que a estratégia de agregação por pares é, ao mesmo tempo, o que viabiliza o caso multiclasse e a principal fonte de imprecisão a ser endereçada.

6. Limitações

Reconhecemos as seguintes limitações, que delimitam o alcance das conclusões. Primeiro, os resultados reportam a média de 100 *holdouts*, mas não acompanham desvio padrão nem testes de significância; diferenças pequenas, como os 0,1 pontos percentuais entre EPCA e MCEPCA em *Bank*, não podem ser afirmadas como reais sem um teste pareado (por exemplo, Wilcoxon por base ou Friedman com pós-teste de Nemenyi para a comparação global entre métodos e bases [Demšar 2006]). Segundo, as densidades foram modeladas como normais; para distribuições assimétricas ou multimodais, essa hipótese pode subestimar a sobreposição, o que é uma explicação plausível para o desempenho do EPCA em *Iris*. Terceiro, a agregação multiclasse por média simples assume implicitamente classes equiprováveis e pode diluir a contribuição de componentes seletivamente discriminantes. Quarto, a avaliação se restringe a dez bases tabulares do UCI e a quatro



(a) *Letter* (26 classes).



(b) *Wine* (3 classes).

Figura 6. Acurácia por número de componentes em mais duas bases multiclasse.

classificadores, limitando a generalização.

7. Conclusão e Trabalhos Futuros

Apresentamos o EPCA, um método supervisionado de seleção de componentes principais que estima o erro de cada componente pela área de sobreposição das densidades de cada classe, usando o coeficiente de Weitzman como aproximação do erro de Bayes. O EPCA elimina duas restrições do MCEPCA, o limite de duas classes e a exigência de covariâncias iguais, e se estende ao caso multiclasse pela média das sobreposições par a par. Nos experimentos, o EPCA equiparou-se ao MCEPCA em problemas binários e, na maioria das bases multiclasse, selecionou componentes mais discriminantes que o PCA clássico e o SPCA. Como trabalho futuro, pretendemos estimar as densidades de forma não paramétrica (por *kernel density estimation*), removendo a hipótese de normalidade; substituir a média das sobreposições par a par por um coeficiente definido diretamente para mais de duas densidades, obtendo estimativa de erro mais precisa em problemas multiclasse; incorporar testes de significância estatística ao protocolo; e medir o custo computacional do método frente às alternativas.

Referências

Barshan, E., Ghodsi, A., Azimifar, Z., and Zolghadri Jahromi, M. (2011). Supervised principal component analysis: Visualization, classification and regression on subspaces

and submanifolds. *Pattern Recognition*, 44(7):1357–1371.

Brownlee, J. (2020). *Probability for Machine Learning*.

chin Ivan Chang, Y. (2026). A survey: Potential dimensionality reduction methods. *arXiv*.

DasGupta, A. (2011). *Probability for Statistics and Machine Learning*. Springer Science.

de Carvalho, T. B. A., Sibaldo, M. A. A., Tsang, I. R., and Cavalcanti, G. D. d. C. (2017). Principal component analysis for supervised learning: a minimum classification error approach. *Journal of Information and Data Management*, 8(2):131–145.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30.

Dhaker, H., Ngom, P., Ibrahimou, B., and Mbodj, M. (2019). Overlap coefficients based on kullback-leibler of two normal densities: Equal means case. *Journal of Mathematics Research*.

Dua, D. and Graff, C. (2019). UCI machine learning repository. `http://archive.ics.uci.edu/ml`.

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*. Academic Press, San Diego, CA, EUA, 2 edition.

Inman, H. F. and Bradley Jr., E. L. (1989). The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Communications in Statistics – Theory and Methods*, 18(10):3851–3874.

Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer Series in Statistics. Springer, New York, 2 edition.

Jose, S., Thomas, S., and Mathew, T. (2019). Interval estimation of the overlapping coefficient of two exponential distributions. *Journal of Statistical Theory and Applications*, 18:26–32.

Linacre, J. M. (1996). Overlapping normal distributions. *Rasch Measurement Transactions*, 10(1).

MCALINN K., T. K. (2026). Optimal hold-out size in cross-validation. *arXiv*.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Vogelstein, J.T., B. E. T. M. (2021). Supervised dimensionality reduction for big data. *Nat Commun*, 12.

Webb, A. R. (2002). *Statistical Pattern Recognition*. John Wiley & Sons, Chichester, Inglaterra, 2 edition.

Weitzman, M. S. (1970). Measures of overlap of income distributions of white and negro families in the united states. Technical Paper 22, U.S. Bureau of the Census, Washington, D.C.