

# EEE928 - Técnicas Clássicas de Reconhecimento de Padrões

## *Metric Learning + KNN*

José Geraldo Fernandes  
Escola de Engenharia  
Universidade Federal de Minas Gerais  
Belo Horizonte, Brasil  
josegeraldof@ufmg.br

Implementou-se um classificador *K-Nearest Neighbours* (KNN) utilizando um aprendizado de métrica para estudar o efeito desse na transformação do espaço de dados e na performance do modelo. Todo o código desenvolvido nesse trabalho está disponível em repositório no GitHub.

### I. GERAÇÃO DE DADOS

Gerou-se dois conjuntos de dados de gaussianas multivariadas para representar duas classes de discriminação. O número de dimensões é  $n = 2$ , as médias das distribuições são  $\mu_1 = [3, 3]^T$  e  $\mu_2 = [3, 4]^T$ , e as variâncias das distribuições são  $\sigma_1^2 = [5, 0.2]^T$ , não há covariância.

A Figura 1 mostra o *plot* dos dados gerados. Note que há sobreposição das classes.

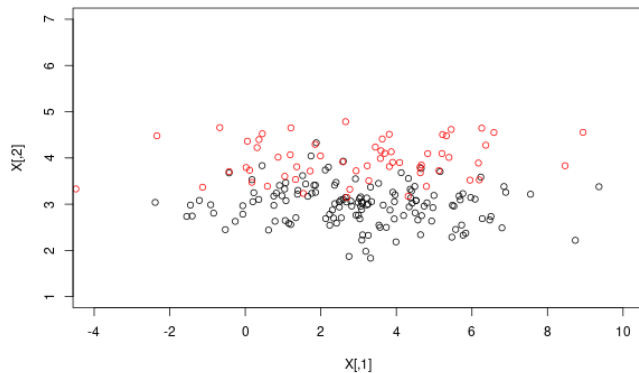


Figura 1. Distribuição dos dados gerados, classe 1 em preto e classe 2 em vermelho.

### II. APRENDIZADO DE MÉTRICA

O problema de sobreposição de classes pode ser caracterizado como uma falta de coesão desses agrupamentos. O aprendizado de métrica, portanto, objetiva minimizar esse efeito buscando maximizar a silhueta média do conjunto de dados.

A silhueta é um método de validação de agrupamentos, calcula-se uma relação entre o quão próximo estão as amostras de um *cluster* e o quão distantes estão essas de *clusters* estranhos, para cada amostra. A acurácia média é o parâmetro utilizado nesse trabalho para avaliar a coesão do conjunto.

Aplicou-se esse método para avaliar o conjunto de dados gerado, encontrou-se uma acurácia média geral de  $silh = 0.11$  e  $silh_1 = 0.16$ ,  $silh_2 = -0.02$  para cada classe. Essa medida varia de  $-1$  a  $+1$ , sendo um valor alto indicador de alta coesão.

A Figura 2 mostra um *plot* de barras para a silhueta calculada de cada amostra de forma ordenada e separada entre classes. Note que, como esperado, a coesão do conjunto é baixa como indica a sobreposição e o baixo valor médio.

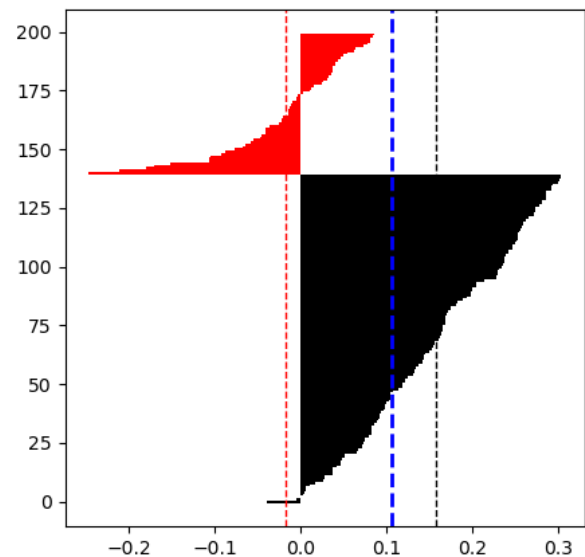


Figura 2. Silhueta ordenada de cada amostra para cada classe, classe 1 em preto e classe 2 em vermelho. As linhas tracejadas mostram a acurácia média do grupo.

Para o aprendizado de métrica, utilizou-se a métrica generalizada de Mahalanobis, como na Equação 1. A matriz  $M$  é o parâmetro de otimização.

$$\delta^2(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T M (\mathbf{x}_i - \mathbf{x}_j) \quad (1)$$

Para o processo de otimização, decompôs-se a matriz na forma  $M = L^T L$ , dessa forma o problema é trabalhado como uma transformação no espaço de dados, da forma  $X^* = LX$ .

Também, utilizou-se a silhueta média geral como função objetivo,  $\text{silh}_{X^*} = \text{silh}(LX, Y)$ . O problema de otimização é definido, portanto, como na Equação 2.

$$L^* = \arg \max_L \text{silh}(LX, Y) \quad (2)$$

Para otimização, foi utilizado o gradiente descendente como na Equação 3, sendo  $lr$  um parâmetro de controle de passo. Inicializou-se a matriz  $L$  a partir de uma distribuição normal uniforme e  $lr = -1$ .

$$\Delta L = L_0 - lr \frac{\partial}{\partial L} \text{silh} \quad (3)$$

A Figura 3 mostra a distribuição dos dados no espaço aprendido e a Figura 4 mostra a silhueta dessas amostras, encontrou-se uma acurácia média geral de  $\text{silh} = 0.43$  e  $\text{silh}_1 = 0.44$ ,  $\text{silh}_2 = 0.40$  para cada classe. Note como a sobreposição é bastante reduzida e os dados são projetados aumentando sua discriminação em ambas dimensões. Isso levanta um interesse futuro em um estudo sobre o efeito de redução de dimensionalidade potencial no aprendizado de métrica.

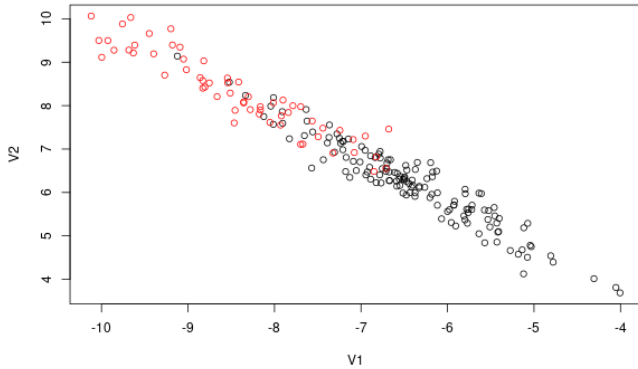


Figura 3. Distribuição dos dados gerados no espaço aprendido, classe 1 em preto e classe 2 em vermelho.

Além disso, também monitorou-se a função objetivo durante o processo. A Figura 5 mostra a evolução nas iterações. O comportamento da curva é característico do método de gradiente descendente.

Outra forma de verificar a coesão dos dados é através da matriz de distâncias. As Figuras 6 e 7 mostram uma

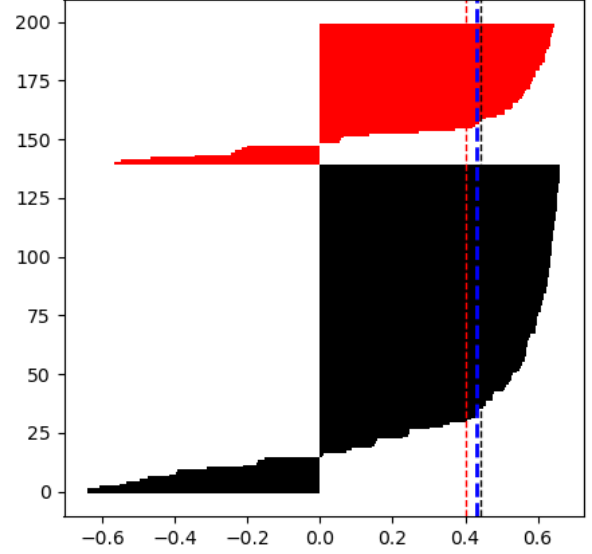


Figura 4. Silhueta ordenada de cada amostra para cada classe no espaço aprendido, classe 1 em preto e classe 2 em vermelho. As linhas tracejadas mostram a acurácia média do grupo.

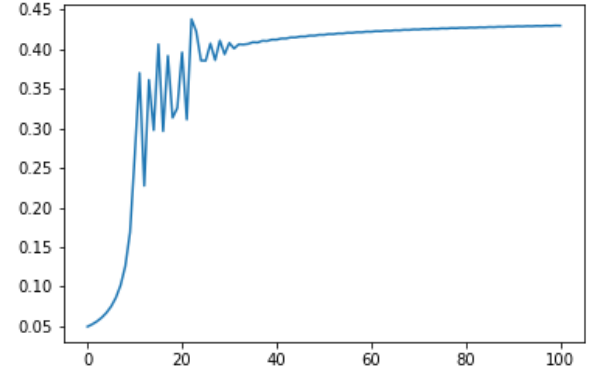


Figura 5. Curva da função objetivo, silhueta média, em iterações.

representação visual dessas matrizes. Note que como as amostras estão ordenadas, um conjunto coeso é percebido com uma diferenciação diagonal característica.

Tanto a avaliação pela silhueta como pela matriz de distâncias mostram o ganho em coesão conseguido pelo aprendizado de métrica.

### III. CLASSIFICADOR

Avaliou-se o ganho do aprendizado de métrica, também, comparando a performance de um classificador KNN.

Para esse experimento, separou-se o conjunto de dados em 10 *folds* para validação cruzada. Apenas o conjunto de treinamento foi utilizado no aprendizado de métrica. Uma

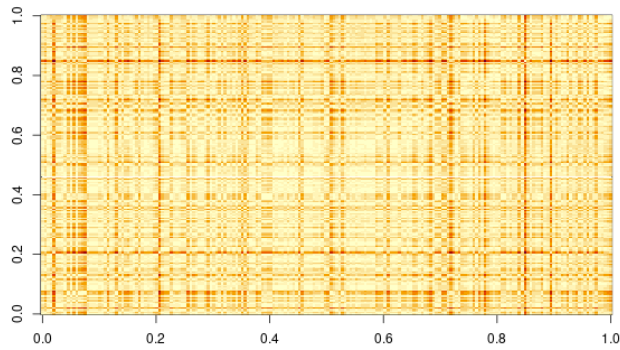


Figura 6. Matriz de distâncias para os dados gerados.

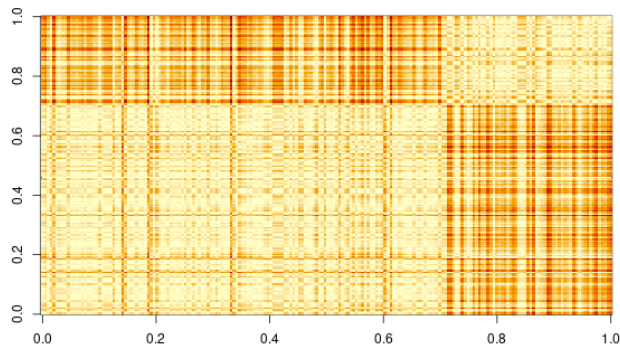


Figura 7. Matriz de distâncias para os dados gerados no espaço aprendido.

varredura simples foi utilizada para selecionar o parâmetro  $n$  de número de vizinhos para o KNN.

A Tabela I mostra a acurácia média do classificador e o número de vizinhos para cada situação.

Dados	$n$	Acurácia média
Original	9	$0.85 \pm 0.06$
Aprendido	13	$0.89 \pm 0.05$

Tabela I

PERFORMANCE DO CLASSIFICADOR EM AMBOS CONJUNTOS DE DADOS,  $\mathbf{X}$  E  $\mathbf{LX}$ .

Diferente das avaliações anteriores, o ganho potencial do aprendizado de métrica não é tão perceptível, principalmente pelo desvio padrão da acurácia, apesar do ganho em média. Isso mostra que o classificador é resistente a falta de coesão no espaço nesse problema, o que torna o esforço adicional de transformação menos necessário.

Segue, portanto, que a validação do método pelo classificador é inconclusiva por sobreposição.