

Projeto AM 2019-2

Francisco de A. T. de Carvalho¹

1 Centro de Informatica-CIn/UFPE
Av. Prof. Luiz Freire, s/n -Cidade Universitaria, CEP 50740-540, Recife-PE, Brasil,
fatc@cin.ufpe.br

1) Considere os dados "Image Segmentation" do site uci machine learning repository

(<http://archive.ics.uci.edu/ml/>). Essa tabela de dados contém 2100 objetos e 7 classes. Os objetos são

descritos por 19 variáveis que podem ser divididas em 2 views: i) Shape view: as primeiras 9 variáveis; ii)

RGB view: as 10 últimas variáveis.

- Execute o algoritmo "Variable-wise kernel fuzzy clustering algorithm with kernelization of the metric" na tabela shape view (2100 objetos e 9 variáveis) e na tabela RGB view (2100 objetos e 10 variáveis), 100 vezes para obter uma partição fuzzy em 7 grupos e selecione o melhor resultado segundo a função objetivo. Para atualizar os protótipos dos grupos use a equação (27). Para atualizar os pesos de relevância das variáveis use a equação (31). Finalmente, para atualizar o grau de pertinência dos objetos nos grupos use a equação (32).
- Para detalhes do algoritmo "Variable-wise kernel fuzzy clustering algorithm with kernelization of the metric" veja as seções 3.1.1 e 3.1.2 do artigo: M.R.P. Ferreira and F.A.T. de Carvalho, Kernel fuzzy c-means with automatic variable weighting, Fuzzy Sets and Systems, 237, 1-46, 2014".
- A partir da partição fuzzy, produza uma partição crisp em 7 grupos.
- calcule o índice de Rand corrigido em relação à partição à priori em 7 classes, para os resultados obtidos com cada view
- calcule o índice de Rand corrigido entre as partições obtidas com cada view
- Observações:
 - Seja $\mathbf{x} = (x_{k1}, \dots, x_{kj}, \dots, x_{kp})$ the k -th object and $\mathbf{v}_i = (v_{i1}, \dots, v_{ij}, \dots, v_{ip})$ the i -th prototype. Use o kernel gaussiano: $\mathcal{K}(x_{kj}, v_{ij}) = \exp \left\{ \frac{-(x_{kj} - v_{ij})^2}{2\sigma_j^2} \right\}$
 - the term $2\sigma_j^2$ is computed as the mean of the 0.1 and 0.9 quantiles of $\|x_{kj} - x_{kl}\|^2 \mid l \neq k$.
 - Parametros: $c = 7$; $m = 1.6$; $T = 150$; $\epsilon = 10^{-10}$;
 - Para o melhor resultado imprimir: i) a matrix de protótipos ii) o vetor de peso de relevância das variáveis iii) a partição crisp (para cada grupo, a lista de objetos), iii) o numero de objetos de cada grupo crisp, iv) 0 indice de Rand corrigido.

- 2) Considere novamente os dados "Image Segmentation" e as 2 views, shape view e RGB view
 - a) Use validação cruzada estratificada "30 times ten fold" para avaliar e comparar os classificadores combinados descritos abaixo. Se necessário, retire do conjunto de aprendizagem, um conjunto de validação para fazer ajuste de parametros e depois treine o modelo novamente com os conjuntos aprendizagem + validação.
 - b) Obtenha uma estimativa pontual e um intervalo de confiança para a taxa de acerto de cada classificador;
 - c) Usar o Wilcoxon signed-ranks test (teste não paramétrico) para comparar os classificadores;

Considere os seguintes classificadores:

- i) Classificador combinado pela regra da soma a partir do classificador bayesiano gaussiano (Um para cada view). Classificador bayesiano gaussiano: considere a seguinte regra de classificação:

afetar o exemplo \mathbf{x}_k à classe ω_j se $P(\omega_j|\mathbf{x}_k) = \max_{i=1}^7 P(\omega_i|\mathbf{x}_k)$ com

$$P(\omega_j|\mathbf{x}_k) = \frac{p(\mathbf{x}_k|\omega_j)P(\omega_j)}{\sum_{r=1}^7 p(\mathbf{x}_k|\omega_r)P(\omega_r)}$$

- a) Use a estimativa de máxima verossimilhança de $P(\omega_i)$
- b) Para cada classe ω_i ($i = 1, \dots, 7$) use a seguinte estimativa de máxima verossimilhança de $p(\mathbf{x}_k|\omega_i) = p(\mathbf{x}_k|\omega_i, \theta_i)$, supondo uma normal multivariada:

$$p(\mathbf{x}_k|\omega_i, \theta_i) = (2\pi)^{-\frac{d}{2}} (|\Sigma^{-1}|)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_k - \mu_i)^T \Sigma^{-1} (\mathbf{x}_k - \mu_i) \right\}, \text{ onde}$$

$$\theta_i = \left(\begin{matrix} \mu_i \\ \Sigma_i \end{matrix} \right), \Sigma_i = \text{diag}(\sigma_i^2, \dots, \sigma_i^2) = \sigma_i^2 \mathbf{I}$$

As estimativas de máxima verossimilhança são: $\mu_i = \frac{1}{n} \sum_{k=1}^n \mathbf{x}_k$,

$$\sigma_i^2 = \frac{1}{d \times n} \sum_{k=1}^n \|\mathbf{x}_k - \mu_i\|^2$$

- c) Classificador combinado pela regra da soma: afetar o exemplo \mathbf{x}_k a classe ω_j se

$$(1 - L)P(\omega_j) + P_{\text{GAUSS,SHAPE}}(\omega_j|\mathbf{x}_k) + P_{\text{GAUSS,RGB}}(\omega_j|\mathbf{x}_k) = \max_{r=1}^7 \left[(1 - L)P(\omega_r) + P_{\text{GAUSS,SHAPE}}(\omega_r|\mathbf{x}_k) + P_{\text{GAUSS,RGB}}(\omega_r|\mathbf{x}_k) \right]$$

- ii) Usar um classificador combinado pela regra da soma a partir do classificador bayesiano baseado em k-vizinhos para fazer a classificação dos dados.
- a) Treine dois classificadores bayesianos baseados em k-vizinhos, um para cada view. Normalize os dados e use a distância Euclidiana para definir a vizinhança. Use conjunto de validação para fixar o o número de vizinhos k .

- b) Regra da soma: afetar o exemplo \mathbf{x}_k a classe ω_j se

$$(1 - L)P(\omega_j) + P_{KVIZ,SHAPE}(\omega_j|\mathbf{x}_k) + P_{KVIZ,RGB}(\omega_j|\mathbf{x}_k) = \max_{r=1}^7 \left[(1 - L)P(\omega_r) + P_{KVIZ,SHAPE}(\omega_r|\mathbf{x}_k) + P_{KVIZ,RGB}(\omega_r|\mathbf{x}_k) \right]$$

com $L = 2$ (DUAS views: SHAPE e RGB)

Observações Finais

- No Relatório e na saída das ferramentas deve estar bem claro:
 - a) como foram organizados os experimentos de tal forma a realizar corretamente a avaliação dos modelos e a comparação entre os mesmos. Fornecer também uma descrição dos dados.
- Data de apresentação e entrega do projeto: **terça-feira** 03/12/2019
- Enviar por email : o programa fonte (se houver), o executável (se houver), os dados e o relatório do projeto
- Tempo de apresentação: 10 minutos (rigoroso).
- PASSAR NA MINHA SALA PARA ASSINAR A ATA DE ENTREGA DO TRABALHO EM 12/06/2019
- ALUNOS DE PÓS-GRADUAÇÃO: O PROJETO DEVE SER REALIZADO COM 3 (TRÊS) ALUNOS.
- ALUNOS DE GRADUAÇÃO: CONSIDERAR 3 GRUPOS, CADA GRUPO COM NO MAXIMO 5 ALUNOS
- ENVIAR PARA fatc@cin.ufpe.br OS INTEGRANTES DE CADA GRUPO, INFORMANDO SE SÃO ALUNOS DE GRADUAÇÃO OU DE PÓS GRADUAÇÃO