

CPE 722 - 1ª série de exercícios - Agrupamentos, Classificadores, Ruído

Obs: A resolução pode ser em grupo. Mas como série visa operar e consolidar os conceitos, é importante que cada um tente entender o problema e encontrar a sua solução.

1 - Os oito elementos à classificar apresentados a seguir são uni-dimensionais:

-3,2 -3,0 -0,7 0,0 0,8 2,0 2,4 2,9

1.1 - Calcule a média m_0 e a dissimilaridade total F_0 do conjunto de elementos.

1.2 - Classifique os elementos usando um classificador k-means cujos três padrões iniciais são $-3,0$ $-0,7$ e $2,0$. Calcule o F_{in} e F_{out} obtidos á partir de suas fórmulas de definição, comprovando que sua soma é igual ao F_0 calculado anteriormente.

1.3 – Para a classificação obtida em 1.2 calcule a dispersão inter-classes das duas classes mais afastadas pelo critério a) do vizinho mais próximo; b) do vizinho mais distante; c) da distância entre os baricentros.

1.4 - A partir do classificador obtido em 1.2 reduza o número de classes de uma unidade usando um algoritmo associativo. Que classes você escolhe para associar, e qual os F_{in} e F_{out} após a associação ?

1.5 – A partir do classificador obtido em 1.2 aumente o número de classes de uma unidade usando um algoritmo dissociativo. Que classe você escolhe para dissociar, e qual os F_{in} e F_{out} após a dissociação ?

1.6 – A partir dos resultados acima é possível concluir algo a respeito do nível do “agrupamento natural ótimo” ?

2 - Considere um conjunto de classes C_i , $i = 1, 2, \dots$ com baricentros $\underline{w}_i$ e raio de similaridade r_0 .

2.1 - Mostre que a equação do plano separador entre duas classes cujos centros são $\underline{w}_j$ e $\underline{w}_k$ pelo critério do padrão mais próximo (critério 1) é

$$2 \underline{x}^t (\underline{w}_j - \underline{w}_k) + |\underline{w}_k|^2 - |\underline{w}_j|^2 = 0$$

2.2 - Mostre que a equação da esfera separadora pelo critério de similaridade mínima r_0 (critério 2) é

$$\underline{x}^t (\underline{x} - 2\underline{w}_k) + |\underline{w}_k|^2 - r_0^2 = 0$$

2.3 - Mostre que o domínio da classe C_k com baricentro $\underline{w}_k$ e raio de similaridade r_0 pelos critérios de centro mais próximo (critério 1) e de similaridade mínima (critério 2) é definido por todo $\underline{x}$ que satisfaz simultaneamente as inequações

$$\begin{aligned} \underline{x}^t (\underline{x} - 2\underline{w}_k) &\leq r_0^2 - |\underline{w}_k|^2 \\ 2\underline{x}^t (\underline{w}_i - \underline{w}_k) &\leq |\underline{w}_i|^2 - |\underline{w}_k|^2 \quad \forall i \neq k \end{aligned}$$

3 - Considere um classificador em que todas as entradas $\underline{x}$ e todos os padrões de classe $\underline{w}_i$ tem módulo unitário. Para este caso mostre que:

a - $u_i = \underline{x}^t \underline{w}_i$ (como nos neurônios das redes feedforward) é uma medida de similaridade entre $\underline{x}$ e $\underline{w}_i$

b - a condição para que uma entrada $\underline{x}$ pertença à uma classe com centro em $\underline{w}$ e raio de similaridade r_0 é que $u_i = \underline{x}^t \underline{w}_i > 1 - (r_0^2 / 2)$.

4 – As figuras abaixo apresentam as projeções das entradas em cada uma das suas 3 dimensões. Estime o número de classes e o desvio padrão do ruído das mesmas em cada dimensão,

