

Atividade Prática 4

Prof. Cristiano Leite de Castro

9 de outubro de 2020

1 TAREFAS PRELIMINARES

1. Ler o Capítulo 12 “Adaptive Neuro-Fuzzy Inference System” do livro texto: Jyh-Shing Roger Jang and Chuen-Tsai Sun. 1996. Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence.
2. Resolver os problemas a seguir.

2 PROBLEMA: CLASSIFICAÇÃO DE PADRÕES

Projete um sistema nebuloso adaptativo (ANFIS) para a resolução dos seguintes problemas de classificação de padrões:

- Breast Cancer Wisconsin (Diagnostic) Data Set, disponível no link:
<https://www.kaggle.com/uciml/breast-cancer-wisconsin-data>
Observação: a coluna ID Number deve ser removida durante a análise pois não traz informação útil para o problema.
- Iris Species Data Set, disponível no link:
<https://www.kaggle.com/uciml/iris>
Observação: este problema tem originalmente 3 classes. Para fins de simplificação,

considere que o objetivo aqui é distinguir a espécie “Iris-setosa” das outras duas espécies existentes. Novamente, a coluna ID deve ser removida durante a análise pois não traz informação útil para o problema.

Para cada problema, divida o conjunto de dados em subconjuntos de treinamento (70%) e teste (30%). Utilize o subconjunto de treinamento para treinamento e seleção de modelos. A seleção de modelos consiste em escolher a combinação mais apropriada de hiperparâmetros a partir de um método de busca em grade via Validação Cruzada com k folds. O espaço de busca de hiperparâmetros compreende o número de regras do ANFIS e o tipo de função pertinência a ser usada no antecedente de cada regra.

O sistema projetado deverá ser do tipo Takagi-Sugeno de ordem 1, ou de ordem 0. Vocês podem utilizar algum mecanismo para inicialização dos parâmetros das funções de pertinência do antecedente como, por exemplo, o algoritmo de agrupamento Fuzzy C-means. Como já visto, ele é útil para encontrar os centros e dispersões iniciais para funções de pertinência do tipo Gaussiana. Após a seleção de modelos, recomenda-se re-treinar o ANFIS com a melhor combinação de hiperparâmetros usando todo o conjunto de treinamento. Utilize este modelo treinado para calcular a Acurácia de classificação para o subconjunto de teste.

No Matlab, o modelo fuzzy Takagi-Sugeno de ordem 1 foi originalmente projetado para resolver problemas de regressão. Para que ele seja capaz de resolver um problema de classificação, vocês devem usar alguma abordagem que adapte a saída do ANFIS. Uma abordagem simples seria modelar o problema normalmente, como se este fosse de regressão e então utilizar um limiar na saída do modelo ($y_{estimado}$), tal como este (limiar de 0.5):

$$\hat{y} = \begin{cases} \text{Classe 1} & \text{se } \hat{y} \geq 0.5 \\ \text{Classe 0} & \text{se } \hat{y} < 0.5 \end{cases} \quad (2.1)$$

Descreva no seu relatório as decisões tomadas e discuta os resultados obtidos. Faça um gráfico com o erro de treino e de teste em cada experimento realizado e discuta o comportamento do erro de treino e teste em função da quantidade de regras do ANFIS.