

Trabalho Computacional III

Prof. Cristiano Leite de Castro

8 de novembro de 2019

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: PREVISÃO DE UMA SÉRIE TEMPORAL

Observações da série temporal de *Mackey-Glass* são produzidas por uma equação diferencial com atrasos de tempo, como segue

$$\dot{x} = \frac{0.2x(t - \Delta)}{1 + x^1(t - \Delta)} - 0.1x(t), \quad (2.1)$$

A equação acima pode ser usada para modelar a dinâmica de produção de células brancas no corpo humano. Pelo fato das taxas de ploriferação destas células envolverem um atraso de tempo, dinâmicas periódicas e caos podem ser verificados. *Mackey and Glass*

sugeriram que flutuações de longo prazo no número de células observadas em certas formas de leucemia apresentam uma dinâmica semelhante à observada por essa equação.

O método de *Runge-Kutta* de quarta ordem foi usado para resolver numericamente a equação acima e os valores da série temporal resultante estão disponíveis no arquivo “mg.mat”.

O objetivo nesta tarefa é usar valores passados da série até o instante t para prever o valor da série em um instante futuro ($t + P$).

Para isso, gere um sistema ANFIS do tipo Takagi-Sugeno de ordem 1 com quatro entradas, correspondendo aos valores $[x(t - 18), x(t - 12), x(t - 6), x(t)]$. A saída do modelo é $ys = x(t + 6)$, i.e., previsão com 6 passos a frente. O número de funções de pertinência associados a cada variável de entrada deve ser igual a 2. Considerando todas as combinações possíveis tem-se então um total de 16 regras. Use funções de pertinência do tipo sino generalizado (*gbell*) e defina valores iniciais razoáveis para os parâmetros de cada uma delas. O sistema ANFIS a ser implementado terá então 104 parâmetros de ajuste, dos quais 24 pertencem aos antecedentes das regras (não-lineares) e 80 pertencem aos consequentes das regras (lineares).

Separe os primeiros 80% dos pontos da série para treinar o sistema ANFIS e teste o desempenho de previsão com os 20% restantes (conjunto de teste). Mostre um gráfico ilustrando os valores previstos e reais em relação aos conjuntos de treinamento e teste. Calcule o MSE (*Mean Squared Error*) produzido para cada um desses conjuntos.

3 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. Você 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 hiperparametros 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):

$$\hat{y} = \begin{cases} \text{Classe 1} & \text{se } \hat{y} \geq 0.5 \\ \text{Classe 0} & \text{se } \hat{y} < 0.5 \end{cases} \quad (3.1)$$

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.