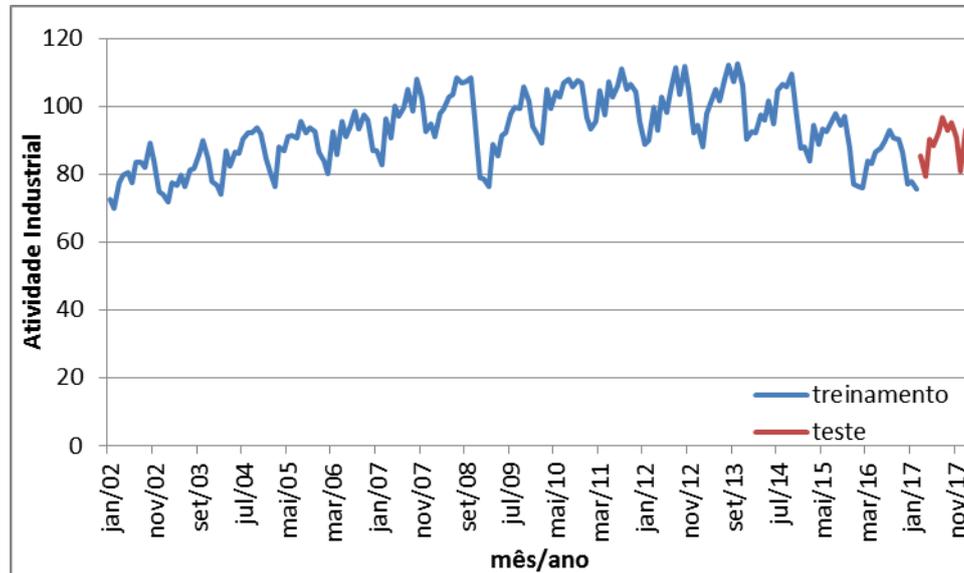


## Previsão de Séries Temporais com Programação Genética (versão #2 atualizada em 18/04/2018)

Este exercício tem o objetivo de utilizar a Programação Genética para gerar um modelo para a previsão de séries temporais. Os 194 dados utilizados são referentes à atividade industrial (mensal) no Brasil de janeiro/2002 a fevereiro/2018 (dados oficiais do Banco Central). A figura a seguir ilustra a série.



Os dados de janeiro/2002 a fevereiro/2017 (182 pontos) devem ser utilizados para a criação do modelo (**TREINAMENTO**) através da PG. Inferindo-se uma possível sazonalidade de  $n$  meses nos dados, um *dataset* deve ser criado contendo colunas com o valor dos  $n$  meses anteriores (isto é,  $n$  variáveis preditoras ou variáveis independentes) como entradas para o modelo, cuja saída é a previsão no mês posterior (atributo meta ou variável dependente). Os casos de fitness são os valores dos 182 meses excluindo-se os  $n$  primeiros meses que entram como variáveis preditoras.

O conjunto de funções é de livre escolha do usuário (podendo criar novas funções) e o conjunto de terminais inclui as  $n$  variáveis preditoras, o atributo meta e, eventualmente, constantes arbitrárias, a critério do usuário. Não há restrição para a profundidade das árvores criadas e evoluídas pela PG, porém a solução final deve ser restrita em até **50** nós. Observe que quanto maior o número de nós, mais complexa será a árvore. Assim, o número de nós deve ser ajustado com parcimônia.

Utilizando o método da Programação Genética, encontrar uma função  $y(t)=f(x(t-1), x(t-2)...x(t-n))$  para o conjunto de pontos dados na planilha, objetivando uma precisão mínima seja de 10% para todos os pontos de treinamento. Utilizar preferencialmente o Lil-GP, porém, outro software similar poderá ser utilizado.

Após o treinamento e obtenção do modelo, simplificá-lo de modo a ficar **humanamente compreensível**. A partir deste modelo, fazer a previsão para os 12 meses subsequentes (março/2017 a fevereiro/2018) para testar a capacidade de previsão do modelo (**TESTE**). Utilizar uma janela deslizante de  $n$  entradas para prever a próxima saída, depois deslizar a janela e incluir a última saída como uma entrada e assim por diante. Com os valores preditos pela PG, calcular o Erro Médio Quadrático (RMSE) entre os valores calculados e os valores reais (fornecidos na planilha).

Preparar um relatório objetivo contendo:

1. Informar os diversos conjuntos parâmetros, funções e terminais que foram testados na PG, e aqueles que efetivamente obtiveram o melhor ajuste.
2. A melhor expressão-S obtida (com identificação, para melhor entendimento).
3. A equação correspondente **já simplificada**.
4. Uma planilha com os pontos  $(x_i, y)$  dados, o valor obtido pela equação encontrada pela PG e o erro percentual para cada ponto, tanto para os dados de treinamento quanto para os dados de teste.
5. Um gráfico, na mesma planilha anterior, com os pontos dados e a função obtida, dentro do intervalo de treinamento especificado.
6. No mesmo gráfico, plotar os pontos de teste com os pontos obtidos da previsão com a PG.
7. Calcular o valor do RMSE para o conjunto de TESTE.
8. Conclusão sobre o uso da PG para a previsão de séries temporais.