Tarefa 1 de Métodos Numéricos I – Teoria de Erros

Nome: Matrícula:

1) Objetivo:

O objetivo dessa tarefa é fazer alguns exercícios e algumas implementações sobre teoria de erros aplicada a assunto de métodos numéricos (correspondente à Unidade 1 da disciplina).

2) Organização:

A tarefa é relativa somente à essa unidade. Cada aluno deve fazê-la individualmente e colocá-la em local definido pelo professor. Códigos devem ser feito em C++ e Linux. Para alguns alunos pode-se fazer em outras linguagens ou sistema operacional opcionalmente, desde que liberado pelo professor da cadeira. Os exercícios devem ser feitos em um editor de textos (tipo WORD ou outro) ou então em papel e escaneados. Depois deve ser gerado um PDF que deve conter as questões resolvidas, junto com os códigos desenvolvidos. Os códigos devem também ser entregues, assim como os executáveis. Executáveis devem incluir todas as bibliotecas usadas. Todos os arquivos, incluindo fontes, executáveis e os exercícios, devem estar juntos em um único arquivo compactado, a ser entregue pelo aluno.

3) O que entregar:

Um único arquivo compactado contendo:
a) Um PDF com todos os exercícios resolvidos.
b) Código fonte das implementações desenvolvidas.
c) Executável das implementações desenvolvidas.

<u>OBS:</u> Recomenda-se que o executável não tenha nada dinâmico, ou seja, que as LIBs sejam estáticas ou todas as DLLs estejam incluídas na distribuição do arquivo.

4) Quando entregar:

No dia e local a ser definido pelo professor da disciplina.
Deverá ser entregue somente por um upload no sistema.
Qualquer atraso na entrega da tarefa não será permitido.

<u>OBS:</u> Não enviar nenhuma tarefa para email do professor!

5) Questões:

Questão 1:

Com relação à conversão binário-decimal e vice-versa pede-se:

a) Faça a conversão binário-decimal de 27 na base 10 para a base 2.
b) Com o resultado do item anterior, faça a conversão de volta para a base 10.
c) Implemente as duas conversões e verifique se os seus resultados estão corretos.

Observação: use as duas maneiras de fazer a conversão de binário para decimal.

Questão 2:

Seja um conjunto de números dados em aritmética de ponto flutuante na base 10, com t=4 e o expoente entre [-5,5]. Pede-se:

a) Qual o menor (m) e o maior (M) número para esse conjunto?
b) O número 100.000 pode ser representado por esse conjunto? Explique.
c) Represente o número 357,26 usando o arredondamento.
d) Represente o número 357,26 usando o truncamento.
e) Supondo que o número 357,26 usando o arredondamento seja igual ao valor exato desse número e que o número 357,26 usando o truncamento seja igual ao valor aproximado, calcule o erro relativo e absoluto desse número.

Questão 3:

Um número em aritmética de ponto flutuante completa é formado por dois fatores que envolvem f_x e g_x. Dito isso, pede-se:

a) Mostre quem seriam f_x e g_x para o número 357,26 usando o mesmo conjunto da questão anterior, ou seja, base 10, com t=4 e o expoente entre [-5,5].
b) Diga quanto valeria os erros absoluto e relativo desse número usando uma equação e inequação, considerando o truncamento.
c) Diga quanto valeria os erros absoluto e relativo desse número usando uma equação e inequação, considerando o arrendodamento simétrico.

Questão 4:

O erro total de um número em aritmética de ponto flutuante é dado pelo erro nas parcelas mais o erro residual de cada operação. Dito isso, pede-se:

a) Diga quanto vale o erro relativo para u = (m+n)w/o supondo que o erro relativo dos números vale $\frac{1}{2} \times 10^{-t+1}$ e usando o arredondamento.
b) Diga quanto vale o erro relativo para o mesmo u, mas agora supondo que os números são representados exatamente e usando o truncamento.
c) Se os valores aproximados para m, n, o e w valem, respectivamente, 10, 20, 30 e 40, calcule quanto valem os erros dos dois itens anteriores.