تمرین شماره ۱ درس محاسبات عددی

سوال ۱:

الف:

دقت کنید طبق فرمول داده شده داریم:

$$|e^x - p_n(x;0)| = \frac{|x|^{n+1}}{(n+1)!}e^c$$

پس برای برقراری رابطه‌ی داده شده باید داشته باشیم:

$$\frac{|x|^{n+1}}{(n+1)!}e^c \leq 10^{-5} \quad \forall x \in [-1,1] \quad c \in [0,x]$$

برای برقراری نامساوی بالا تنها لازم است مقدار n را برای حالتی که جمله‌ی e^c بیشینه است محاسبه کنیم. زیرا اگر n انتخاب شده برای بیشینه‌ی جمله‌ی بالا نامساوی را برقرار کند، برای تمامی x ها نیز این نامساوی برقرار است. حال داریم:

$$\underset{x}{\operatorname{argmax}} \frac{|x|^{n+1}}{(n+1)!}e^c = 1 \Rightarrow \frac{e^c}{(n+1)!} \leq 10^{-5}$$

از طرفی چون مقدار c در بازه‌ی $[0,x]$ قرار دارد و تابع exponential صعودی است، برای بیشینه کردن عبارت باید c را برابر با x یعنی 1 قرار دهیم و داشته باشیم:

$$e \times 10^5 \leq (n+1)!$$

که عبارت بالا برای $n \geq 8$ برقرار است.

ب:

دقت کنید تابع entire exponential است، که به این معنا است بسط تیلور آن برای تمامی اعداد در صفحه‌ی مختلط برقرار است و می‌دانیم بسط e^x به صورت مقابل است:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

حال با جایگذاری عبارت بالا در انتگرال داده شده خواهیم داشت:

$$\int_0^1 \frac{e^x - 1}{x}dx = \int_0^1 1 + \frac{x}{2!} + \frac{x^2}{3!} + \dots dx$$

با گرفتن انتگرال از جملات $\sum_{n=0}^{\infty} \frac{x^n}{(n+1)!}$ در بازه‌ی $[0,1]$ خواهیم داشت:

$$\int_0^1 \frac{e^x - 1}{x}dx = 1 + \frac{1}{2 \times 2!} + \frac{1}{3 \times 3!} + \dots = \sum_{k=1}^{\infty} \frac{1}{k \times k!}$$

حال می‌دانیم در نهایت $10^{-6} \times \frac{1}{2}$ خطای گرد کردن داریم پس خطای truncate ما باید برابر با $10^{-6} \times \frac{1}{2}$ باشد که به صورت مقابل محاسبه می‌شود:
* دقت کنید برای محاسبه‌ی خطا از رابطه‌ی بخش الف استفاده می‌کنیم:

$$\int_0^1 \frac{x^n e^c}{(n+1)!} = \frac{e^c}{(n+1)(n+1)!} \leq \frac{e}{(n+1)(n+1)!} \leq \frac{1}{2}10^{-6} \Rightarrow n \in [9, \infty]$$

پس حال 8 جمله‌ی اول را با دقت 7 رقم اعشار جمع می‌زنیم و در نهایت حاصل را به 6 رقم اعشار گرد میکنیم:

$$\int_0^1 \frac{e^x - 1}{x} dx \approx 1.000000 + 0.2500000 + \cdots = 1.3179018 \approx 1.317902$$

سوال ۲:

الف:

لازم است برای این بخش 2 حالت را در نظر بگیریم:

$$a_t = 0$$

با فرض 0 بودن a_t داریم:

$$fl(x) = \overline{1.a_1 a_2 \dots a_{t-1}} 2^e$$

در این حالت داریم $fl(x) \leq x$ پس باید عبارت مقابل را اثبات کنیم:

$$2^e(\frac{x}{2^e} - \frac{fl(x)}{2^e}) \leq 2^{e-t} \equiv \frac{x}{2^e} - \frac{fl(x)}{2^e} \leq 2^{-t} \equiv \frac{x}{2^e} \leq \frac{fl(x)}{2^e} + 2^{-t} = \overline{1.a_1 a_2 \dots a_{t-1} 1}$$

حال دقت کنید طبق فرض ما برای $\frac{x}{2^e}$ داشتیم $a_t = 0$ و حال نشان داده‌ایم که بیت t ام $\frac{fl}{2^e} + 2^{-t}$ برابر با 1 است و نامساوی بالا اثبات می‌شود.

$$a_t = 1$$

با فرض 1 بودن a_t داریم:

$$fl(x) = (\overline{1.a_1 a_2 \dots a_{t-1}} + 2^{1-t})2^e$$

در این حالت داریم $x \leq fl(x)$ پس باید عبارت مقابل را اثبات کنیم:

$$-2^{e-t} \leq 2^e(\frac{x}{2^e} - \frac{fl(x)}{2^e}) \equiv \frac{fl(x)}{2^e} - \frac{x}{2^e} \leq 2^{-t} \equiv \frac{fl(x)}{2^e} \leq 2^{-t} + \frac{x}{2^e}$$

حال داریم:

$$2^{-t} + \frac{x}{2^e} = 2^{-t} + \overline{1.a_1a_2\ldots a_{t-1}1a_{t-2}\ldots} = \overline{1.a_1a_2\ldots a_{t-1}0a_{t-2}\ldots} + 2^{1-t}$$

حال به صورت بدیهی میدانیم:

$$\overline{1.a_1a_2\ldots a_{t-1}} + 2^{1-t} \leq \overline{1.a_1a_2\ldots a_{t-1}0a_{t-2}\ldots} + 2^{1-t}$$

و حکم بالا اثبات میشود.

ب:

دقت کنید داریم:

$$\overline{1.a_1a_2\ldots} \geq 1 \Rightarrow x \geq 2^e$$

حال طبق نابرابری بالا داریم:

$$\frac{x - fl(x)}{x} \leq \frac{x - fl(x)}{2^e}$$

حال در بخش قبل اثبات کردیم

$$\frac{x - fl(x)}{2^e} \leq 2^{-t}$$

پس در نهایت اثبات می‌شود:

$$\frac{x - fl(x)}{x} \leq \frac{x - fl(x)}{2^e} \leq 2^{-t}$$

سوال ۳:

الف:

روش اول:

دقت کنید وقتی x به سمت 0 میل می‌کند تقریب آن نیز یعنی $\bar{x}$ به سمت 0 میل می‌کند که باعث می‌شود $cos(\bar{x})$ به 1 میل کند. این باعث میشود در صورت خطای تفریق داشته باشیم و برای رفع آن از اتحاد $1 - cos(2x) = 2sin^2(x)$ استفاده می‌کنیم و تابع را به صورت مقابل بازنویسی می‌کنیم:

$$\frac{1 - cos(x)}{x^2} = \frac{2sin^2(\frac{x}{2})}{x^2}$$

همانطور که مشاهده می‌کنید خطای تفریق 2 عدد نزدیک از هم رفع شده است.

روش دوم:

دقت کنید وقتی x به سمت 0 میل می‌کند تقریب آن نیز یعنی $\bar{x}$ به سمت 0 میل می‌کند که باعث می‌شود $cos(\bar{x})$ به 1 میل کند. این باعث میشود در صورت خطای تفریق داشته باشیم و برای رفع آن از بسط تیلور $cos(\bar{x})$ استفاده میکنیم.

$$cos(\bar{x}) = 1 - \frac{x^2}{4!} + \frac{x^4}{6!} - \dots$$

حال با جایگذاری رابطه‌ی بالا داریم

$$f(x) = \frac{1}{4!} - \frac{x^2}{6!} + \frac{x^4}{8!} - \dots$$

همانطور که مشاهده میکنید به دلیل متناوب بودن سری تیلور بالا با استفاده از Remainder Theorem می‌توانیم دقت را تا حد دلخواه کنترل کنیم و از طرفی دیگر خطای تفریق را نداریم.

ب:

در این بخش نیز شاهد خطای تفریق هستیم و برای رفع آن از اتحاد مقابل استفاده می‌کنیم:

$$a^3 - b^3 = (a-b)(a^2 + ab + b^2)$$

با استفاده از اتحاد بالا داریم:

$$x = (\sqrt[3]{x+1} - 1)((x+1)^{\frac{2}{3}} + \sqrt[3]{x+1} + 1) \Rightarrow f(x) = \frac{x}{(x+1)^{\frac{2}{3}} + \sqrt[3]{x+1} + 1}$$

همانطور که دوباره مشاهده می‌شود خطای تفریق دیگر مشکلی برای ما ایجاد نخواهد کرد.

ج:

در این بخش نیز برای رفع خطای تفریق در صورت از اتحاد مزدوج استفاده می‌کنیم:

$$\frac{\sqrt{4+x} - 2}{x} = \frac{\sqrt{4+x} - 2}{x} \times \frac{\sqrt{4+x} + 2}{\sqrt{4+x} + 2} = \frac{1}{\sqrt{4+x} + 2}$$

همانطور که مشاهده می‌شود خطای تفریق دیگر مشکلی ایجاد نمی‌کند.

سوال ۴:

الف:

روش اول:

هنگامی که x بسیار بزرگ می‌شود، به دنبال آن $\frac{1}{x}$ کوچک می‌شود و محاسبه‌ی عبارت $\sqrt{1+\frac{1}{x}} - 1$ با از دست دادن دقت رقم اعشار مواجه می‌شود. برای رفع این مشکل می‌توانیم

تابع را به صورت مقابل بازنویسی کنیم:

$$\sqrt{1+\frac{1}{x}}-1=\sqrt{1+\frac{1}{x}}-1\times\sqrt{1+\frac{1}{x}}+1=\frac{\frac{1}{x}}{\sqrt{1+\frac{1}{x}}+1}=\frac{1}{\sqrt{x}(\sqrt{x+1}+\sqrt{x})}$$

روش دوم:

دقت کنید محاسبه کردن رابطه‌ی داده شده این اجازه را به ما نمی‌دهد تا مقدار داده را با دقت دلخواه محاسبه کنیم، برای این کار نیاز داریم بسط تیلور عبارت داده شده را بنویسیم و سپس از Remainder Theorem استفاده کنیم که می‌گوید خطای بسط تیلوری متناوب حداکثر برابر با آخرین جمله‌ای است که محاسبه نشده است.

$$\sqrt{1+x}=1+\frac{x}{2}-\frac{x^2}{8}+\dots$$

دقت کنید ROC سری بالا اطراف 0 است و حال با جایگذاری $\frac{1}{x}$ در عبارت بالا داریم:

$$\sqrt{1+\frac{1}{x}}-1=\frac{1}{2x}-\frac{1}{8x^2}+\dots$$

دقت کنید چون x بسیار بزرگ است $\frac{1}{x}$ طرفای 0 خواهد بود و مشکلی برای ROC به وجود نخواهد آمد.

حال اگر عبارت بالا را تا جمله‌ی n ام حساب کنیم خطای ما برابر خواهد بود با:

$$E_n \le X_{n+1}$$

ب:

روش اول:

$$xf(x)=\frac{x}{\sqrt{x}(\sqrt{x+1}+\sqrt{x})}=\frac{\sqrt{x}}{\sqrt{x+1}+\sqrt{x}}=\frac{1}{\sqrt{1+\frac{1}{x}}+1}$$

حال داریم:

$$\lim_{x\to+\infty}\frac{1}{x}=0\Rightarrow\lim_{x\to+\infty}\frac{1}{\sqrt{1+\frac{1}{x}}+1}=\frac{1}{\sqrt{1}+1}=\frac{1}{2}$$

روش دوم:

با ضرب x در سری تیلور بالا داریم:

$$xf(x) = \frac{1}{2} - \frac{1}{8x} + \frac{1}{16x^2} - \frac{5}{128x^3} + \Omega(x^{-4})$$

حال با میل دادن x به بینهایت خواهیم داشت:

$$\lim_{x\to+\infty} xf(x) = \frac{1}{2}$$

سوال ۵:

الف:

رابطه‌ی بازگشتی را به صورت مقابل بدست می‌آوریم:

$$x^2 = \frac{5}{2}x - 1 = 0 \Rightarrow r_1 = \frac{1}{2}, r_2 = 2$$

حال با جایگذاری مقادیر داده شده و حل دستگاه خطی داریم:

$$x_n = c_1(2)^n + c_2(\frac{1}{2})^n$$

$$\begin{cases} 2c_1 + \frac{1}{2}c_2 = a \\ 4c_1 + \frac{1}{4}c_2 = \frac{a}{2} \end{cases} \quad \Rightarrow c_2 = 2a, c_1 = 0$$

پس رابطه‌ی بازگشتی به صورت مقابل است:

$$x_n = 2a(\frac{1}{2})^n$$

ب:

دقت کنید چون $\frac{1}{3}$ را نمی‌توان در سیستم IEEE به صورت دقیق ذخیره کرد، همواره مقداری خطا خواهیم داشت. برای بررسی رشد نسبی خطا می‌توانیم فرض کنیم اعداد را تا 5 رقم اعشار با قاعده‌ی گرد کردن می‌توانیم نشان دهیم. با فرض انجام شده رابطه‌ی بازگشتی به صورت مقابل بازنویسی می‌شود:

$$P_1 = 0.33333, P_2 = 0.16667 \Rightarrow \begin{cases} 2C_1 + \frac{1}{2}C_2 = 0.33333 \\ 4C_1 + \frac{1}{4}C_2 = 0.16667 \end{cases}$$

با حل دستگاه بالا به مقادیر مقابل برای C_1 و $C2$ می‌رسیم:

$$C_1 = \frac{1}{600000}, C_2 = \frac{49999}{75000} \Rightarrow \bar{P}_n = \frac{1}{600000}2^n + \frac{49999}{75000}(\frac{1}{2})^n$$

حال می‌توانیم رابطه‌ی بازگشتی اصلی را از رابطه‌ی جدید کم کنیم تا خطای مطلق بدست آید:

$$e_n = \bar{P}_n - P_n = \frac{1}{600000}2^n - \frac{1}{75000}(\frac{1}{2})^n$$

پس خطای مطلق به صورت نمایی با پایه‌ی 2 رشد می‌کند و برای پیدا کردن ضریب خطای نسبی، کافی است عبارت بالا را بر مقدار واقعی رابطه‌ی بازگشتی تقسیم کنیم:

$$\delta_n = \frac{\bar{P}_n - P_n}{P_n} = \frac{\frac{1}{600000}2^n - \frac{1}{75000}(\frac{1}{2})^n}{\frac{2}{3}(\frac{1}{2})^n} = \frac{1}{40000}4^n - \frac{1}{50000}$$

پس خطای نسبی به صورت نمایی با ضریب 4 رشد می‌کند و نتایج بالا را با کد مقابل تایید می‌کنیم.

```python
n = []
n.append(1/3)
n.append(1/6)

prev_answer = 1 - n[1]/(1/6)

for i in range(2,1000):
    n.append(5/2 * n[i - 1] - n[i - 2])
    if prev_answer != 0:
        print((1 - n[i] / (2/3 * 1/2**i))/prev_answer)
    prev_answer = 1 - n[i]/(2/3 * 1/2**i)
```

برای بررسی نتیجه‌ی بدست آمده از تکه کد بالا استفاده می‌کنیم.

```
1.0000000000000007
1.0000000000000027
1.0000000000000107
1.0000000000000426
1.0000000000001705
1.0000000000006821
1.0000000000027285
1.000000000010914
1.0000000000436557
1.000000000174623
1.000000000698492
1.0000000027939677
1.000000011175871
1.000000044703483
1.0000001788139237
```

```
1.0000007152555668
1.0000028610202207
1.0000114440481414
1.0000457756687062
1.0001830942935612
1.0007322431047108
1.0029268292682927
1.0116731517509727
1.0461538461538462
1.1764705882352942
1.6
2.5
3.4
3.823529411764706
3.953846153846154
3.9883268482490273
3.9970731707317073
3.9992677568952892
3.9998169057064388
3.9999542243312938
3.999988555951859
3.9999971389797793
3.9999992847444332
3.9999998211860763
3.999999955296517
3.999999988824129
3.9999999972060323
3.999999999301508
3.999999999825377
3.9999999999563443
3.999999999989086
3.9999999999972715
3.999999999999318
3.9999999999998295
3.9999999999999574
3.9999999999999893
3.9999999999999973
3.999999999999999
4.0
```

همانطور که مشاهده می‌شود تقسیم خطای نسبی در مرحله‌ی $n+1$ بر خطای نسبی در مرحله‌ی n پس از چندین Iteration به 4 همگرا می‌شود که نشان می‌دهد رشد خطای نسبی به صورت نمایی و با ضریب 4 است که در توافق با نتایج بدست آمده در بخش تئوری است.

ج:

دقت کنید این بار a برابر با 1 است که برخلاف $\frac{1}{3}$ به صورت دقیق قابل نمایش در فرمت IEEE است.

$$\frac{5}{2}x_{n-1} - x_{n-2} = 2x_{n-1} + \frac{x_{n-1}}{2} - x_{n-2}$$

که از ابتدا داریم $x_{n-2} = 2x_{n-1}$ به همین دلیل نتیجه‌ی نهایی برابر خواهد بود با:

$$\frac{5}{2}x_{n-1} - x_{n-2} = \frac{x_{n-1}}{2}$$

که به دلیل اینکه فرمت در IEEE نمایش داده شده است می‌توان با کاهش exponent به اندازه‌ی یکی به صورت دقیق نشان داد. در نتیجه خطا برابر با 0 خواهد بود که باعث می‌شود خطای نسبی نیز برابر با 0 شود.

```python
n = []
n.append(1)
n.append(1/2)
prev_answer = 1 - n[1]/(1/6)
for i in range(2,1000):
    n.append(5/2 * n[i - 1] - n[i - 2])
    if prev_answer != 0:
        print((1 - n[i] / (1/2**i))/prev_answer)
    prev_answer = 1 - n[i]/(1/2**i)
```

نتیجه‌ی اجرای کد بالا به صورت مقابل است:

```
-0.0
```

دقت کنید هر بار prev_answer برابر با 0 است و به همین دلیل تنها یکبار 0 در کنسول چاپ شده است.

سوال ۶:

```python
import math
from decimal import Decimal

print("Calculating e^(-30) using built in function exp: " + str(
    math.exp(-30)))


x = -30
iterations = 1000

answer = 0

```

```python
#left ro right
for i in range(iterations + 1):
    answer += Decimal(x**i) / Decimal(math.factorial(i))
print("Adding from left to right: " + str(answer))

print("Difference from left to right " + str(answer - Decimal(math.exp(-30))))

answer = 0

#right to left
for i in range(iterations, -1, -1):
    answer += Decimal(x**i) / Decimal(math.factorial(i))

print("Adding from right to left " + str(answer))


print("Difference from right to left " + str(answer - Decimal(math.exp(-30))))
```

نتیجه‌ی اجرای کد بالا به صورت مقابل است:

```
Calculating e^(-30) using built in function exp: 9.357622968840175e-14
Adding from left to right: 9.364788142833340100782216574E-14
Difference from left to right 7.165173993165284164921624157E-17
Adding from right to left 9.3593186E-14
Difference from right to left 1.695631159825183382705050157E-17
```

همانطور که از نتایج بالا مشخص است هنگامی که عبارت e^x را با استفاده از بسط تیلور محاسبه می‌کنیم و جواب آن را برای تعداد محدودی جمله محاسبه می‌کنیم، شروع محاسبه از چپ یا از راست نتایج مختلفی را به ما می‌دهد.

همانطور که مشاهده می‌شود اختلاف جواب واقعی با جوابی که از راست به جمع زدن شروع کرده‌ایم کمتر از حالتی است که از چپ جمع زده‌ایم، که به معنای دقیق‌تر بودن نتیجه‌ی جمع از راست به چپ است.

دلیل این امر این است که زمانی که اعداد بزرگ را ابتدا باهم جمع می‌زنیم تاثیر اعداد کوچکتر کمتر شده و به چشم نمی‌آیند و دقت جمع کاهش می‌یابد، ولی هنگامی که ابتدا جملات کوچک را باهم جمع می‌کنیم حاصل آن‌ها تاثیر بیشتری نسبت به تک تک آن‌ها دارد و جملات کمتری نادیده گرفته می‌شوند که باعث می‌شود نتیجه دقیق‌تر شود.
همانطور که بالاتر توضیح داده شد جمع کردن عبارت از راست به چپ که باعث می‌شود اعداد کوچکتر تاثیر بیشتری داشته باشند نتیجه‌ی دقیق‌تری را حاصل می‌شود که خروجی کد بالا بیانگر این حقیقت می‌باشد.