Data Assignment #5  
20229007 Sungyoung Ryu (BTM539) AI for Business Management

**LSA(latent semantic analysis):**

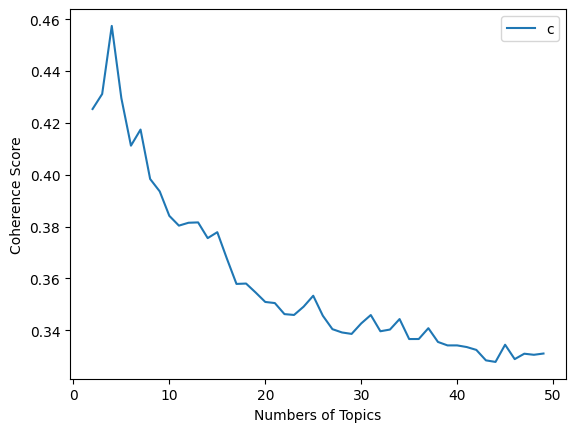
# Scikit-Learn: optimal number of topics(num\_topics) by 'explained\_variance\_ratio\_.sum()'

I could easily figure out 50 of topics is inefficient to find the optimal numbers, because topics are overlapped in same keywords and documents, observes 6,799 columns from pp\_topictermMtx DataFrame. Also we draw the elbow chart; to decide the optimal numbers of explained\_variance\_ratio\_.sum() and there is no dramatically changes (some points got upside down) in topics controlling over pp\_LSA\_sklm.explained\_variance\_ratio\_.sum()

차트이(가) 표시된 사진

자동 생성된 설명  
Perhaps, adjusting min\_df(minimum document frequency) and max\_df(maximum document frequency) values using the CountVectorizer is effective, but require iteration and error to find the optimal values. (Due to computing power of Colab Free version, stopped exploration)

# Gensim: optimal number of coherence\_values



Simply, I can observe the elbow curve in the plot coherence\_values in the coherence score (y axis) from given code by professor. We can expected the peak between 0 to 10 would be the optimal point of topics, in order to iterate, bot find the optimal topic number were 4 at the end. – After then, I used daily limitance of Colab computing, then asked me to use the premium version and no longer explore with free version.

**Individual Project Proposal over UN SDGs 17 for Final term project:**

**Goals #13 Climate Change:** Take urgent action to combat climate change and its impacts

13-2. Integrate climate change measures into national policies, strategies and planning or 13-3. Improve education, awareness-raising and human and institutional capacity on climate change mitigation, adaptation, impact reduction and early warning

This semester is the 2nd half of my Professional Master’s Program(Master of Enterpreneur and Innovation). Once I entered into K-School, my business item tackled sustainability on eco-friendly packaging problem – especially focused aseptic carton (juice box) which is not recycled as much we are doing sorting wastes as well due to lack of recycling plants. Comparing to annual production, only 3% of aseptic carton is being recycled in 2020. It has significant gap with paper(more than 100%) or milk carton(16%). Although I switched my graduation project into a combination of blockchain and AI on digital contents market currently, but still personally targeted on climate action and climate technology in deeply.

So if I could do, I want to challenge the SDGs #13 on climate change sector with programming. Methodology is to be determined. But at least I want to contribute the real deal on BTM539.