Problem 1 (10pt): Pearson Correlation Coefficient

We introduce Pearson Correlation Coefficient in the class. Given two sequences of numbers $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$, the Pearson Correlation Coefficient (PCC) is computed as

$$
PCC = \sum_{i=1}^{n} \frac{x_i - \bar{x}}{\sqrt{\sum_{j=1}^{n}(x_j - \bar{x})^2}} \frac{y_i - \bar{y}}{\sqrt{\sum_{j=1}^{n}(y_j - \bar{y})^2}}
$$

where

$$
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i
$$

Show that $-1 \leq PCC \leq 1$ regardless of the values of $x$s and $y$s.

Problem 2 (10pt): Collaborative Filtering as Information Retrieval

Describe your approach of using a document retrieval system for collaborative filtering. This should include (i) how to represent each training user as a text document, (ii) how to represent each test user as a text query, and (iii) how to make the recommendation of objects given the results returned by the document retrieval system. Furthermore, describe an approach of extending pseudo relevance feedback to collaborative filtering, and discuss the Pros and Cons of using a document retrieval system for collaborative filtering.