Косинусное сходство и мягкое сходство.

И то, и другое – это мера сходства двух векторов, представляющих некоторые массивы данных. Используется для соотнесения содержащихся в этих массивах объёмов информации, определения связей между ними, например при анализе естественного языка.

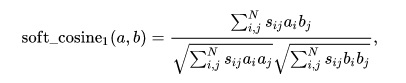
Косинусное сходство рассчитывается исходя из скалярного произведения векторов A и B по формуле:



Соответственно: 1 – полное совпадение; о – полное отсутствие связи; -1 – полная противоположность.

Мягкое косинусное подобие отличает от обычного тем, что оно учитывает сходство между содержащимися в векторах данными (между парами данных в них), тогда когда в случае обычного косинусного сходства вектора рассматриваются как независимые.

Мягкое рассчитывается по формуле:



где Sij – сходство между i, j. Если Sij = 0, т.е. сходства нет, мягкое сходство превращается в обычное косинусное.