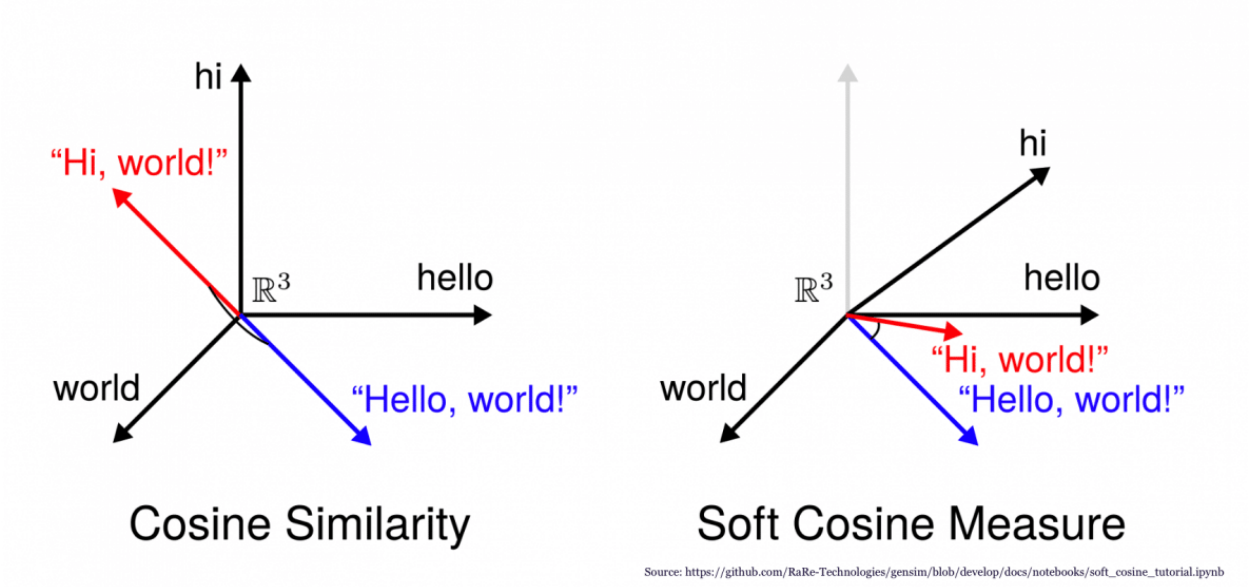
Одной из основных задач ML – является обработка естественного языка, в т.ч. обнаружение сходств слов. Обычно для решения этой задачи используется «Косинусная мера» между двумя объектами (словами), представленными в виде векторов. Но такой метод имеет недостаток: объекты рассматриваются как независимые. Например, если обратиться к картинке из лекции:



Можно заметить, что в первом случае все слова расположены одинаково относительно друг друга. Во втором случае, слова Hi и Hello расположены ближе друг к другу, что является правильным – ведь слова имеют практически один и тот же смысл. Это и есть «Мягкая косинусная мера», т.е. учитывает схожесть «Базовых» векторов (слов), которые образуют векторное пространство. Для вычисления родства слов можно использовать какой-либо словарь синонимов, либо же метод Левенштейна. Метод позволяет вычислить минимальное число односимвольных операций для преобразования одного слова в другое.