

## Mode 3 (sentence embeddings):

*Use case:* learn pairwise similarity from collections of similar objects, e.g. sentence similarity.

*Data format:* each line is a collection of similar sentences:

sent1\_word1 sent1\_word2 ... <tab> sent2\_word1 sent2\_word2 ...

*Training:*

- Each sentence is represented as a bag of features (words or n-grams). They are embedded to predict sentence similarity
- Similar sentence pairs are taken from the collections
- Dissimilar sentence pairs are sampled at random