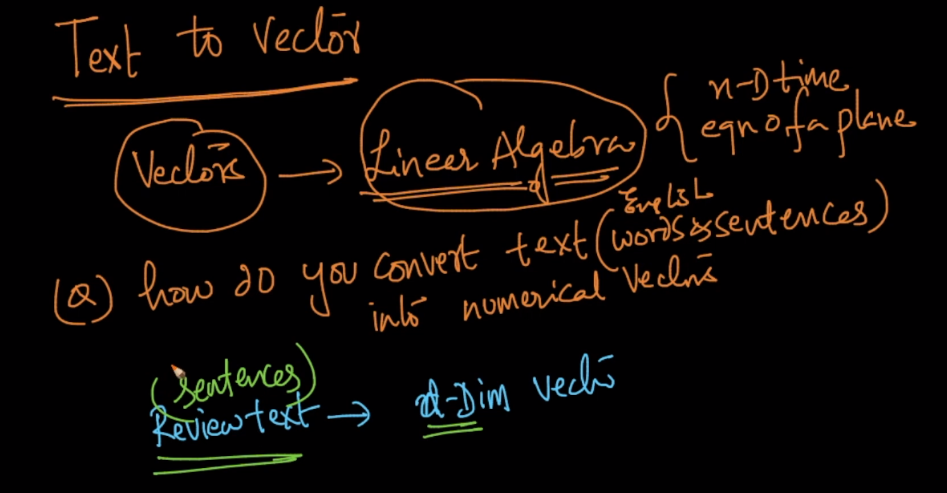
**Why we convert text to vector:**

By converting text to vector, we can leverage methamatical power to do classification or anaylsis.

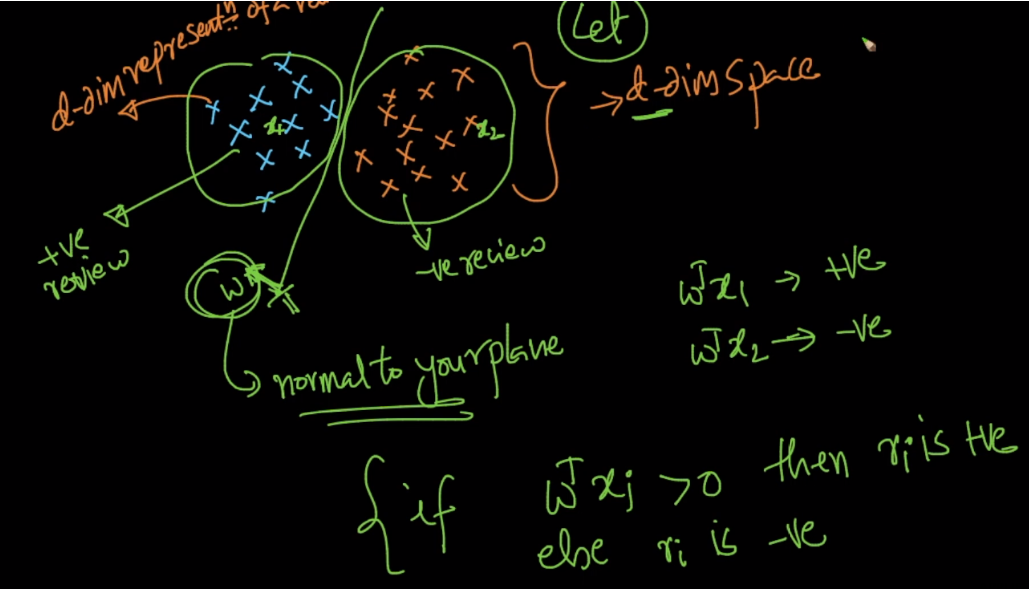
Generally, we are going to convert the text into a vector representation, where each dimension of the vector corresponds to a word, and its value maps in some way to the frequency or importance of the word in text chunk. The transformation can be performed using Bow, idf ,tfidf etc. From there you have a representation of the document in a continuous high-dimensional space and can apply clustering and classification algorithms

We convert each sentence to a d-dim vector



After creating vectors of each sentences we may find that vectors corresponding to positive sentences will be accumulated together and vectors corresponding to negative sentences will be accumulated together.

How we can separate them ( +ve or –ve ), we’ll draw a hyperplane and calculate the direction of normal vector (w), so all the points that are in direction of w (mean wT . x1 is +ve) then it belong to positive sentence’ category, and if it’s –ve it belongs to negative sentence’ category.



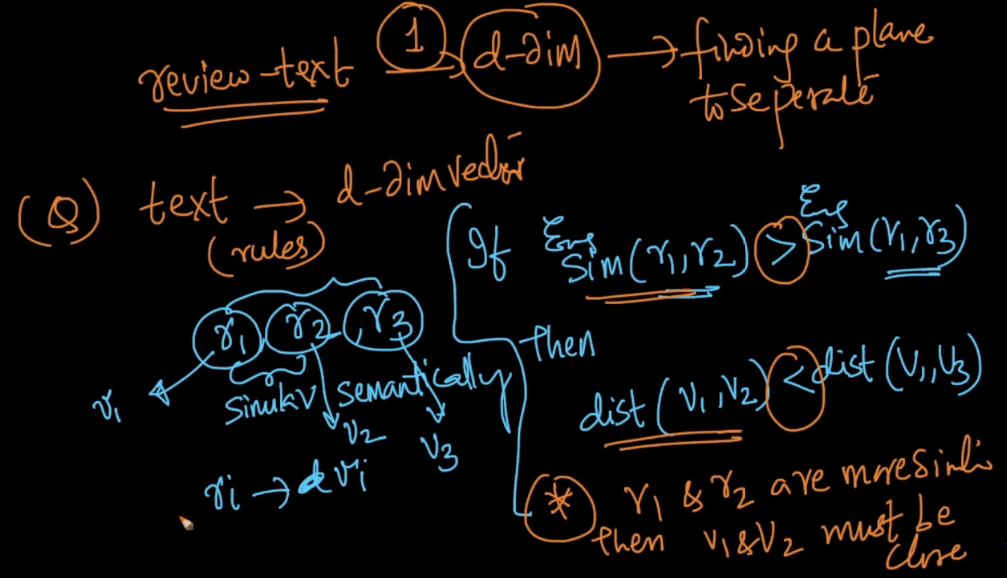
**Rules:**

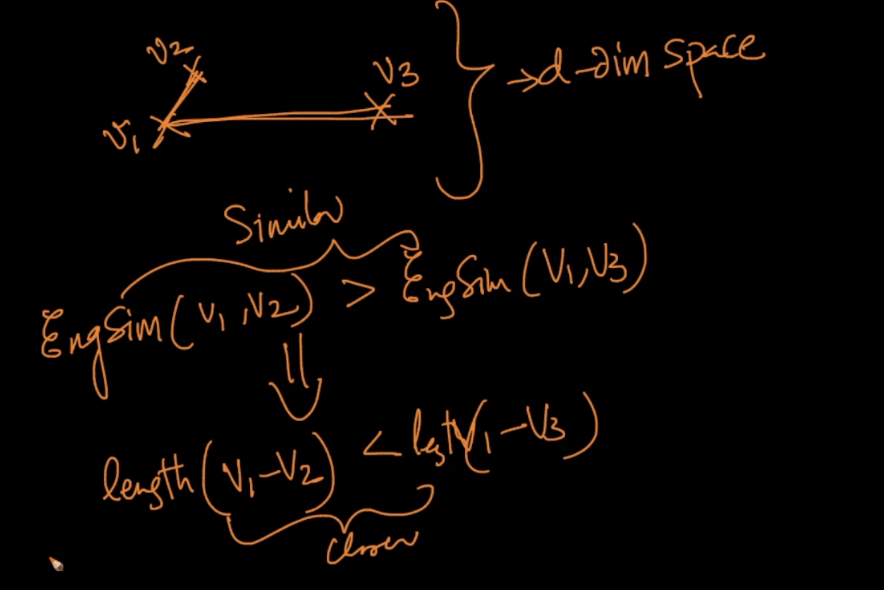
Let’s say we have 3 sentences r1, r2, r3 and their respective vectors v1, v2, v3.

If r1 and r2 are more similar than r1 and r3.

Then distance between v1 and v2 will be much less than distance between v1 and v3.

**Conclusion**: if two sentences are more similar then distance between their vectors will be much close.





There are various techniques to convert text to d-dimension vector as shown in below figure.

