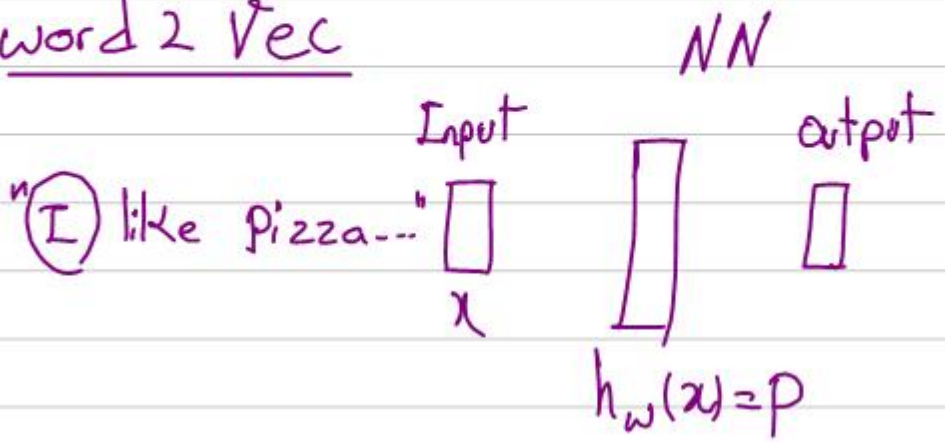
TF / IDF & word2 vec

feature

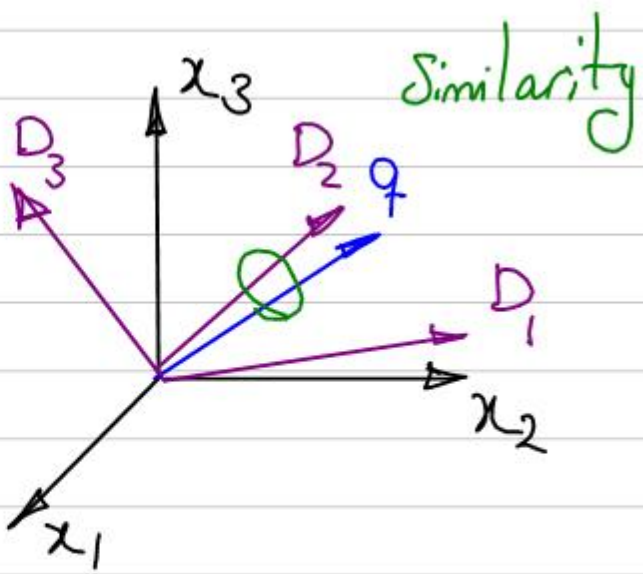
	I	like	hate	...	Predict
Doc1	1	1	0		5
Doc2	1	0	1		1

مثال على

word2 Vec



Vector Space model



Query: جملة

D1

D2

D3

TF / IDF

Sparsity ← One-hot

↓

(a/the) frequent Tokens ← BOW

Sim = TF * IDF ↓

Term Frequency = Coont(w, D) ← TF/IDF

Inverse Doument frequency = $\log\left(\frac{M+1}{df(w)}\right)$

q: "How Can I change my RAM"

---

q: "Nike Shoes" D1: "Puma Nike"

V: {Nike, Puma, shoes, ...} D2: "Nike Shoes Nike" *

D3: "Shoes Puma Puma Shoes Niko shoes Puma"

q = (1, 0, 1) D2 = (2, 0, 1) فرض

D مجموعه: M = 500

IDF(Nike) = $\log \frac{501}{150} = 0.5$ df(Nike) = 150

IDF(puma) = 0.5 df(Puma) = 150

IDf(shoes) = $\log \frac{501}{300} = 0.2$ df(Shoes) = 300

D1 = (1*0.5, 1*0.5, 0*0.2) Sim(q, D1) = 0.5 + 0 + 0 = 0.5

D2 = (2*0.5, 0*0.5, 1*0.2) → Sim(q, D2) = 1 + 0 + 0.2 = 1.2 ←

D3 = (1*0.5, 3*0.5, 3*0.2) Sim(q, D3) = 0.5 + 0 + 0.6 = 1.1