

# First Executions Report

A summary of 3 different experiments is described below

## \* Experiment #1

Vocab  $\rightarrow$  100 queries + 42'900 clicked products

Dim.  $\rightarrow$  128  $\sim$  756 different category

Training Samples: 1'327'707

In the second Exp Dim increased to 300, result got better but yet not acceptable

## \* Not a significant accuracy observed : (

- Seems like the vocab size is so large with respect to Dim. and queries. (100)

Solutions  $\rightarrow$  Dim  $\uparrow$ , Vocab Size  $\downarrow$

Solutions  $\rightarrow$  Dim  $\downarrow$ , Vocab Size  $\uparrow$

- Seems like the vocab size is so large with respect to Dim. and queries. (100)

## \* Not a significant accuracy observed : (

### #3 Exp $\rightarrow$ Only learning Categories.

$\sim$  740 classes

$\rightarrow$  300 Dim vectors.

Everything else remained unchanged from last two Exps.

Results: \* Similar queries observed to be close to each other.

مثلاً ماشین اصلاح و ریش تراش

\* But adjacent categories to each query is lacking of accuracy.

همیشه نزدیکترین فاصله را به ریش تراش می‌دهد!

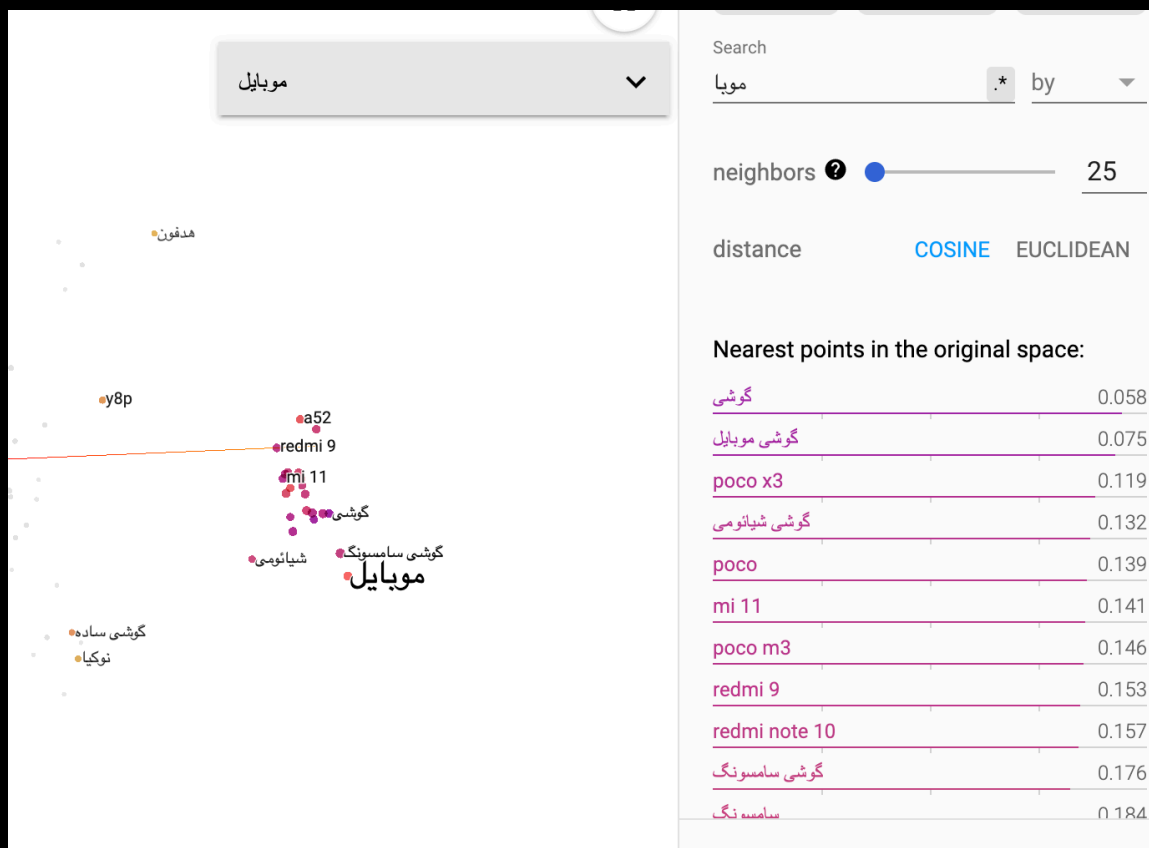
همیشه نزدیکترین فاصله را به ماشین اصلاح می‌دهد!

\* But adjacent categories to each query is lacking of accuracy.

Result evaluation methods were not used yet, but for a better intuition, a few embedding space results related to experiment 3, which categories were to be predicted, are illustrated.

## Acceptable examples

closeness of different categories and queries related to mobile phone



موتور برق

Search

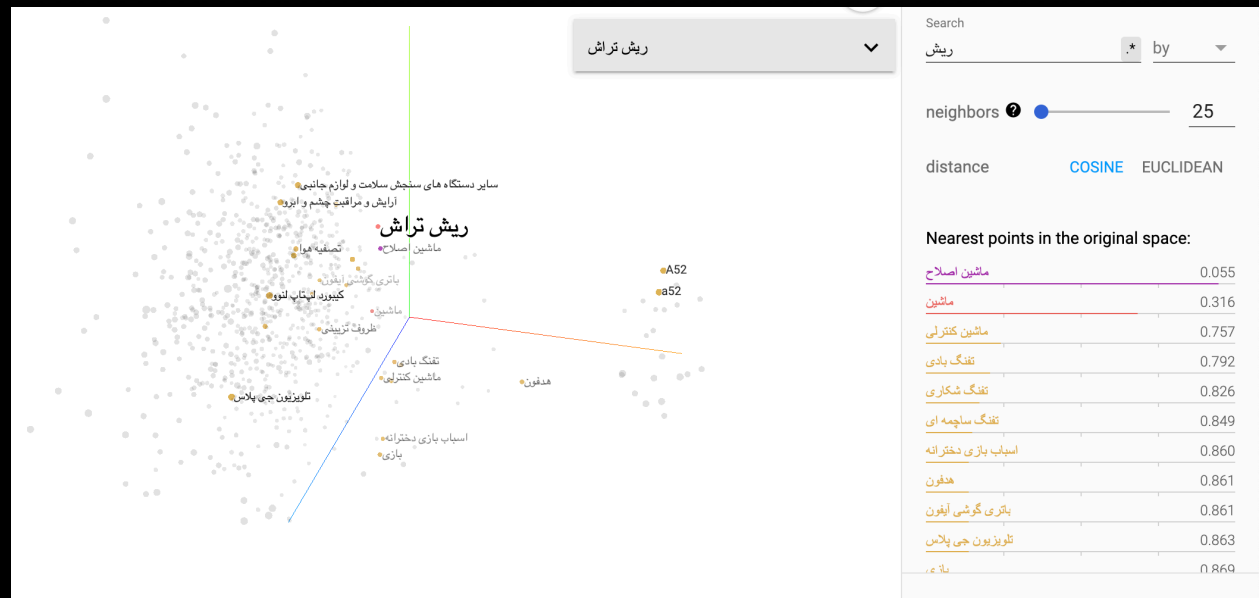
موتور برق

neighbors 25

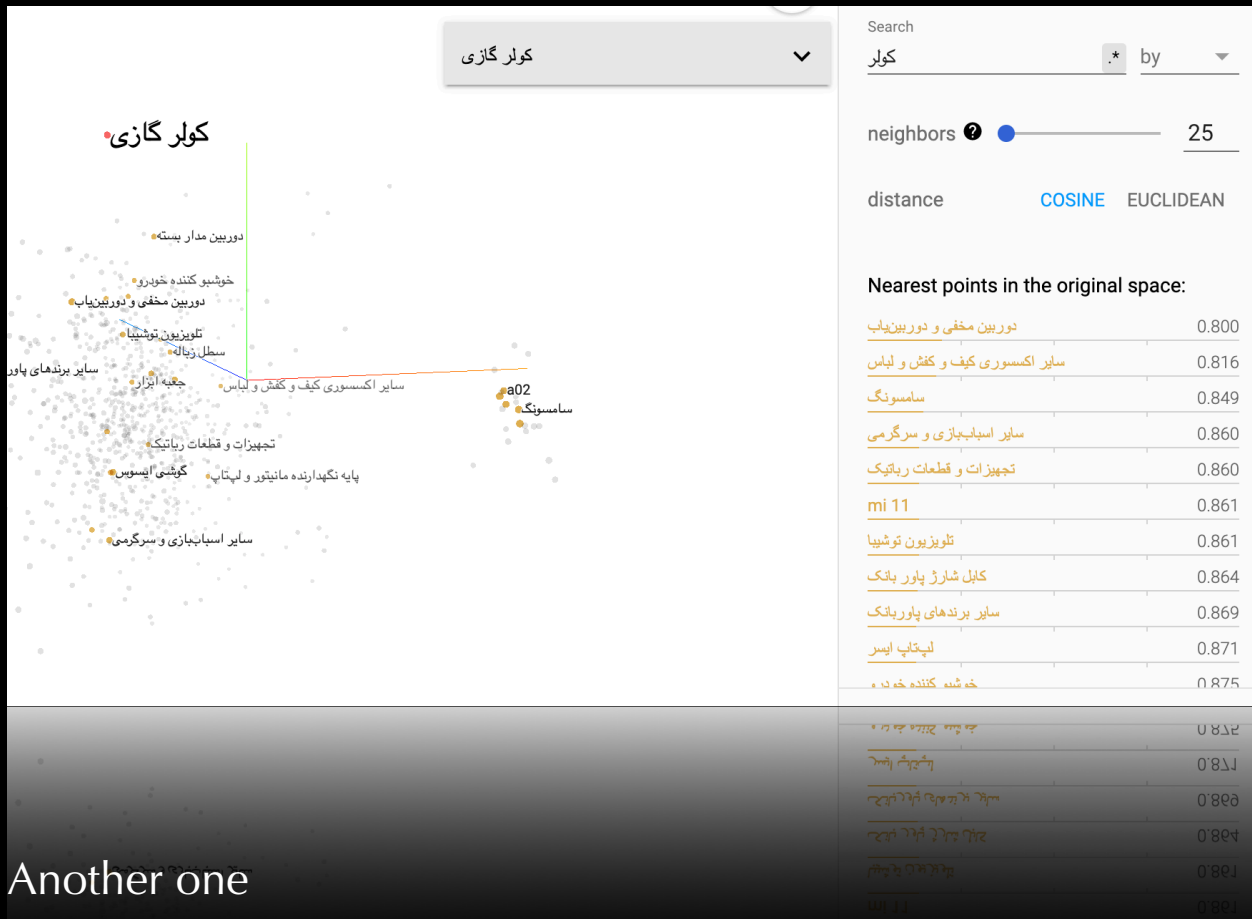
distance COSINE EUCLIDEAN

Nearest points in the original space:

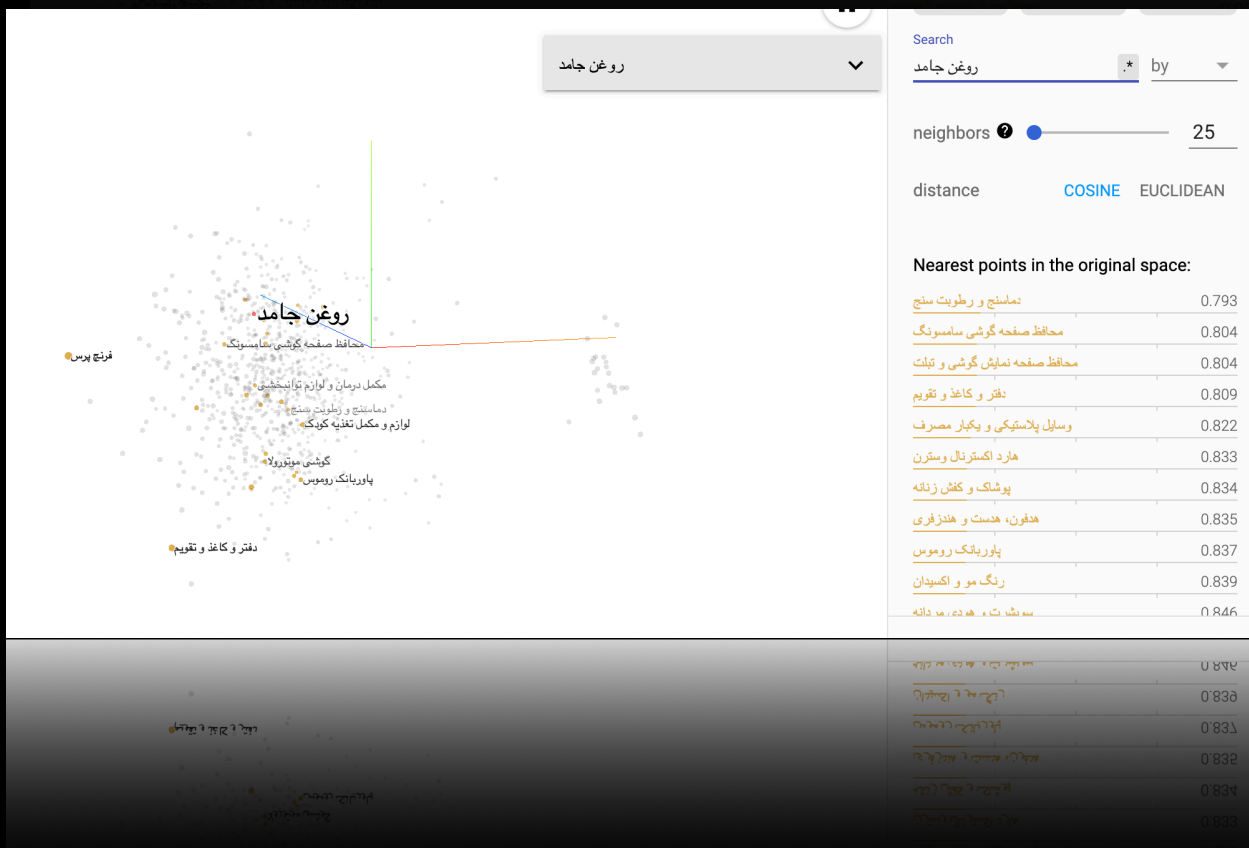
Point Name	Distance
دستگاه جوش	0.195
مینی فرز	0.223
دریل	0.242
اسباب بازی دخترانه	0.765
بازی	0.801
کیف	0.816
هارد دیسک اینترنتال	0.818
باربی	0.821
کیف زنانه	0.826
تلفنک ساجمه ای	0.844
شوک	0.845
چرخ	0.846
بازیچه	0.849
چرخ	0.852
بازی	0.853
بازیچه	0.878
بازی	0.878
بازی	0.878
بازی	0.878



## Unacceptable examples



# Another one



## Uncertainties

- Don't we need to assume a few training samples which its  $X$  is the `ad_id`?
- Do we need to use the "results" section in the search data set?
- Should we pay attention to the "page number" of any search query?