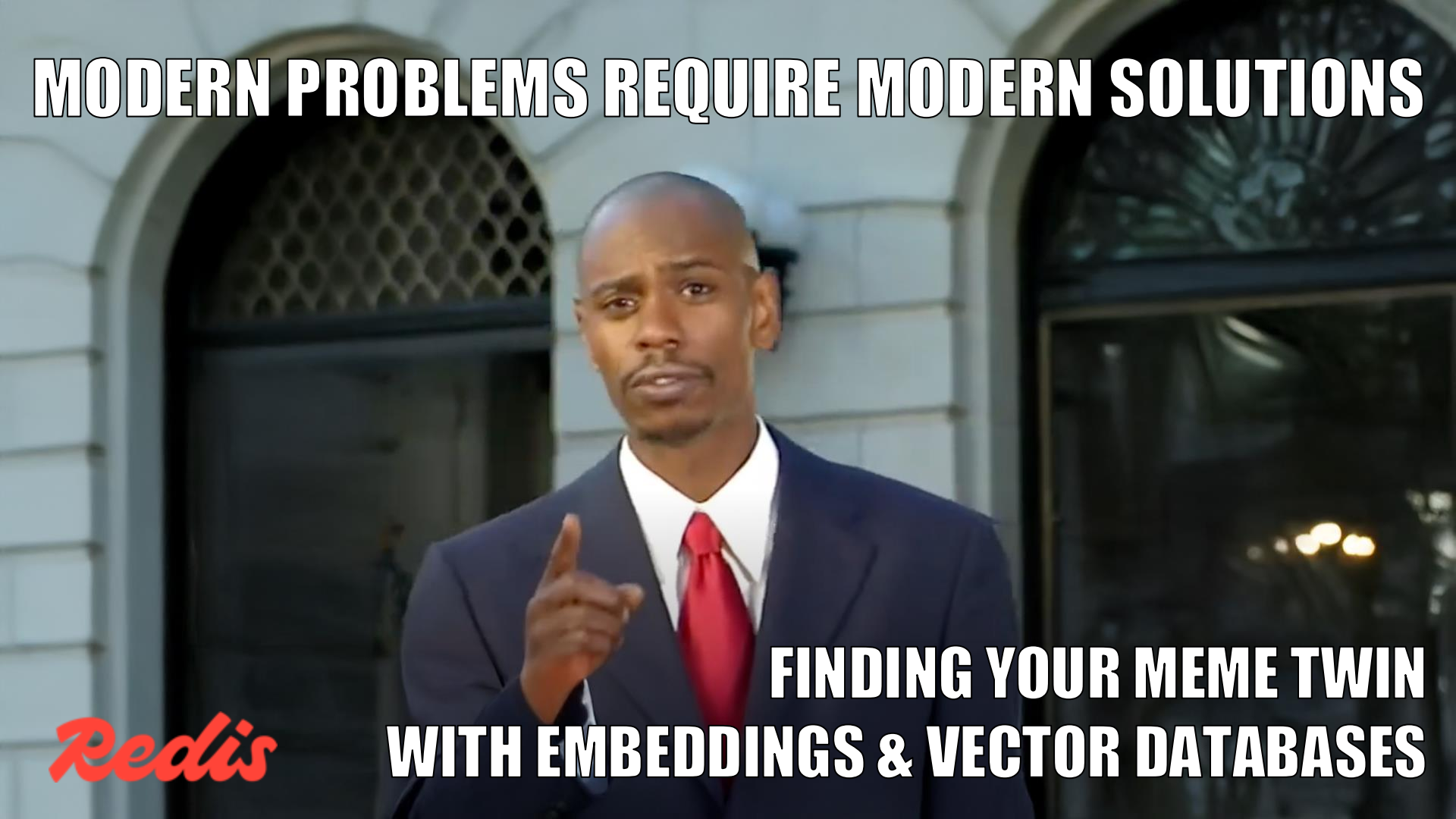

MODERN PROBLEMS REQUIRE MODERN SOLUTIONS
FINDING YOUR MEME TWIN
WITH EMBEDDINGS & VECTOR DATABASES
Redis

Guy Royse

Developer Advocate

Redis

@guy.dev

@guyroyse

github.com/guyroyse

guy.dev

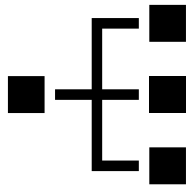

I DON'T ALWAYS PRETEND TO BE A DATA SCIENTIST
BUT WHEN I DO, I DO IT ON STAGE


I'M SOMETHING OF A DATA SCIENTIST MYSELF


```
[  0.86 -0.72  0.25
  -0.93 -0.13 -0.64
   0.47  0.39  0.78
  -0.58 -0.41  0.63  ]
```

What's an embedding?

Using embeddings



Making embeddings

Code & demo


AM I THE ONLY ONE AROUND HERE WHO KNOWS
WHAT AN EMBEDDING IS?
Redis

Structured data

Numbers

Strings

Locations

Unstructured data

Text

Images

Audio

STORING ALL YOUR COMPANY'S STRUCTURED DATA IN SQLITE

STORING ALL YOUR COMPANY'S STRUCTURED DATA IN ACCESS

STORING ALL YOUR COMPANY'S STRUCTURED DATA IN EXCEL

DON'T NEED TO DEBUG YOUR CODE IF IT DOESN'T COMPILE

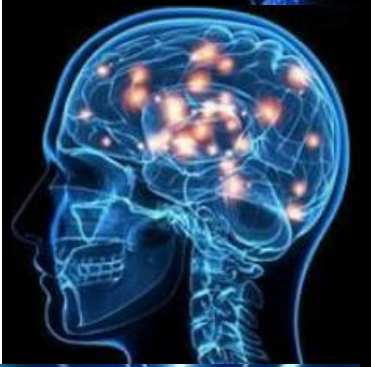
Summarize → Tokenize → Embed

$$\begin{bmatrix} 0.86 & -0.72 & 0.25 \\ -0.93 & -0.13 & -0.64 \\ 0.47 & 0.39 & 0.78 \\ -0.58 & -0.41 & 0.63 \end{bmatrix}$$

Transform → Convolute → Embed

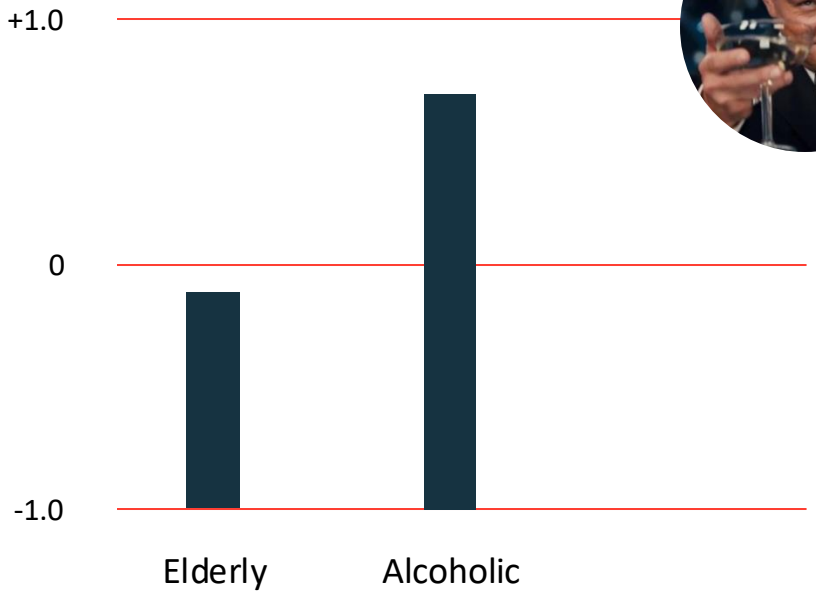
$$\begin{bmatrix} 0.72 & 0.86 & -0.39 \\ -0.63 & 0.95 & -0.21 \\ -0.86 & 0.61 & 0.45 \\ -0.79 & -0.03 & 0.70 \end{bmatrix}$$

HAVING
UNSTRUCTURED
DATA

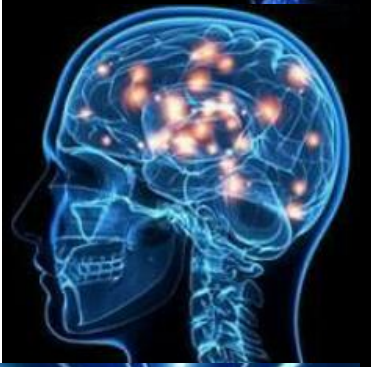


HAVING
STRUCTURED
DATA

USING AI TO GET
FEATURES
FROM
UNSTRUCTURED
DATA

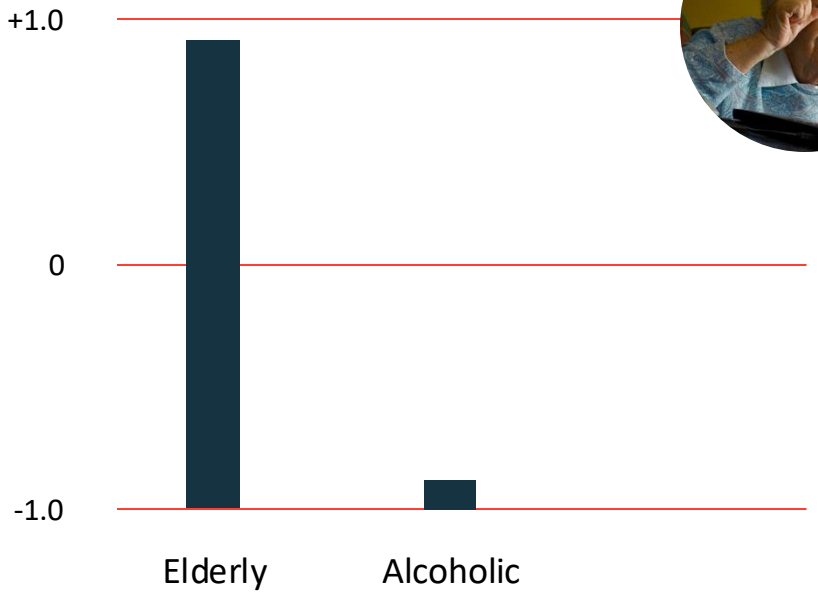


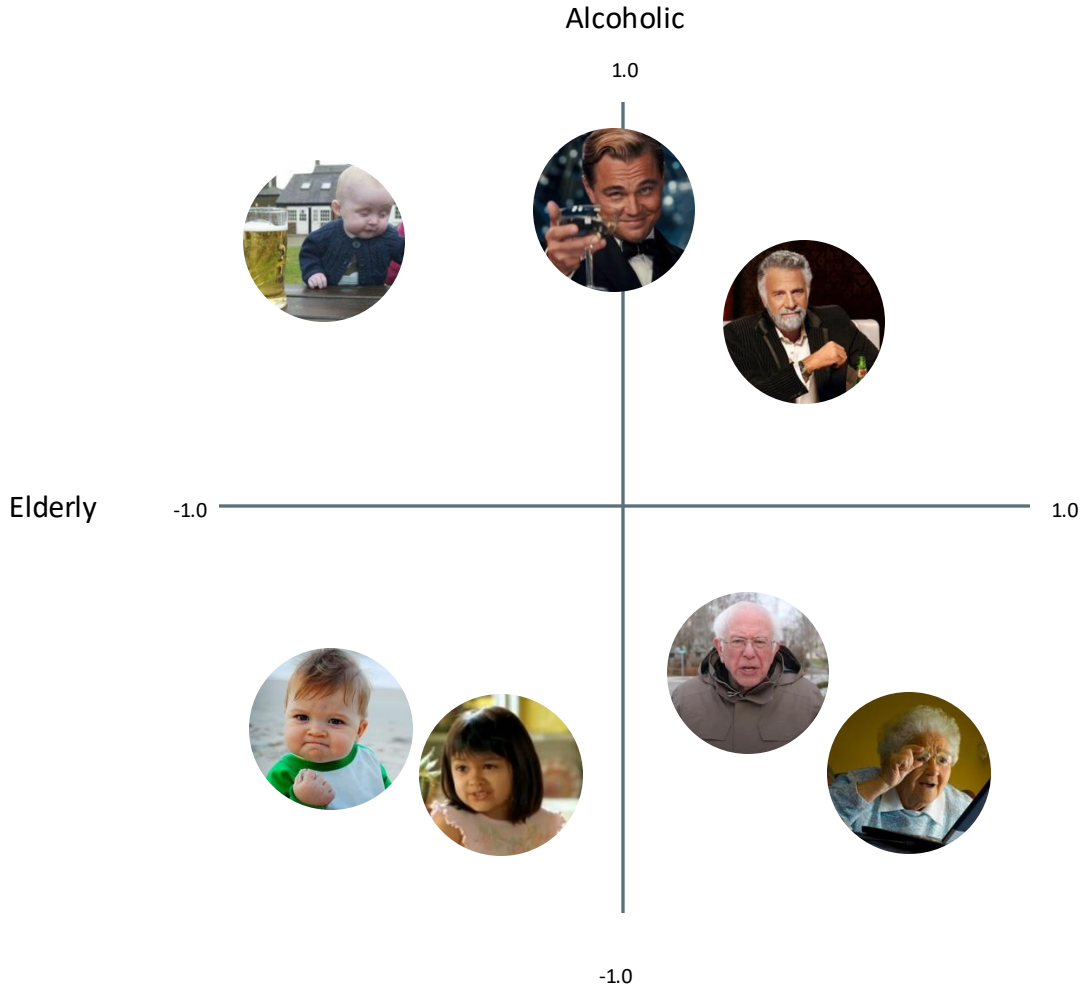
HAVING
UNSTRUCTURED
DATA



HAVING
STRUCTURED
DATA

USING AI TO GET
FEATURES
FROM
UNSTRUCTURED
DATA




Alcoholic
1.0
Elderly
-1.0
1.0
-1.0

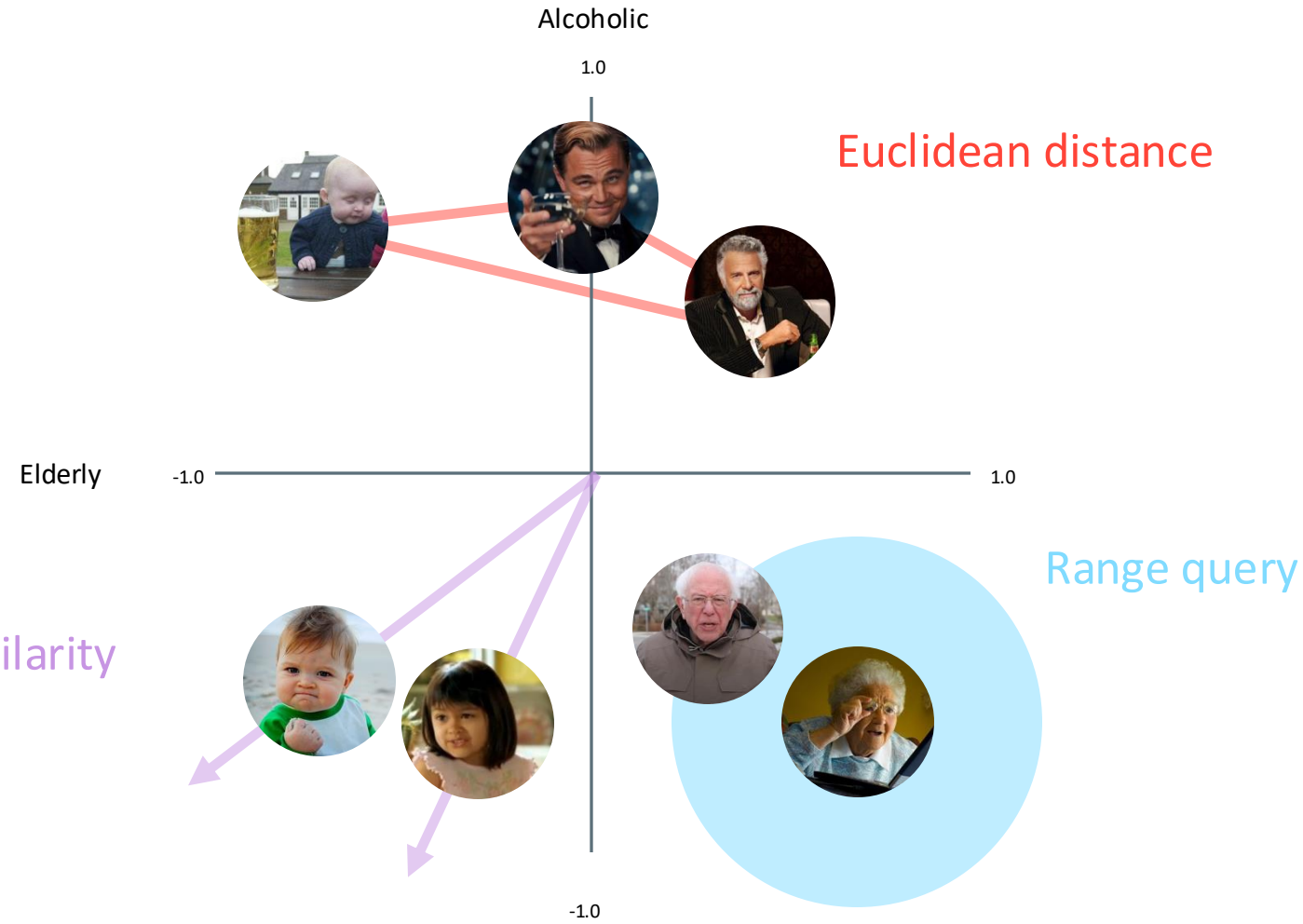
[0.72 0.86 -0.39
-0.63 0.95 -0.21
-0.86 0.61 0.45
-0.79 -0.03 0.70]

[0.86 -0.72 0.25
-0.93 -0.13 -0.64
0.47 0.39 0.78
-0.58 -0.41 0.63]

[-0.96 0.52 -0.77
0.68 -0.95 0.34
0.82 -0.29 0.17
-0.43 -0.61 0.79]

[0.89 -0.76 0.23
-0.90 -0.03 -0.68
0.45 0.39 0.68
-0.42 -0.40 0.57]

Redis


Alcoholic
1.0
Elderly
-1.0
1.0
-1.0
Euclidean distance
Range query
Cosine similarity

Flat

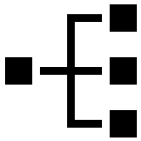
- A simple brute-force search
- Perfect accuracy
- Runs into scaling limitations

HNSW

- Hierarchical navigable small world
- Creates layers of graphs in a hierarchy
- Approximate accuracy
- Tunable to balance accuracy, speed, and space

SEMANTIC SEARCH

DON'T NEED TO DEBUG YOUR CODE IF IT DOESN'T COMPILE



[
0.89 -0.76 0.23
-0.90 -0.03 -0.68
0.45 0.39 0.68
-0.42 -0.40 0.57
]

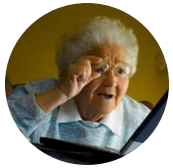
DELETION IS 100% FILE COMPRESSION

TYPESCRIPT IS JUST JAVASCRIPT FOR C# DEVS CHANGE MY MIND

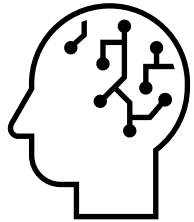
CODE DOESN'T COMPILE. COMPILES IT AGAIN.

SEMANTIC CACHING

Can you explain to me what a dank meme is?

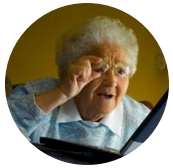


[0.89 -0.76 0.23
-0.90 -0.03 -0.68
0.45 0.39 0.68
-0.42 -0.40 0.57]



RETRIEVAL AUGMENTED GENERATION

Can you explain to me what a dank meme is?



[
0.89 -0.76 0.23
-0.90 -0.03 -0.68
0.45 0.39 0.68
-0.42 -0.40 0.57
]

Overused memes are dank

Dank memes are the dankest

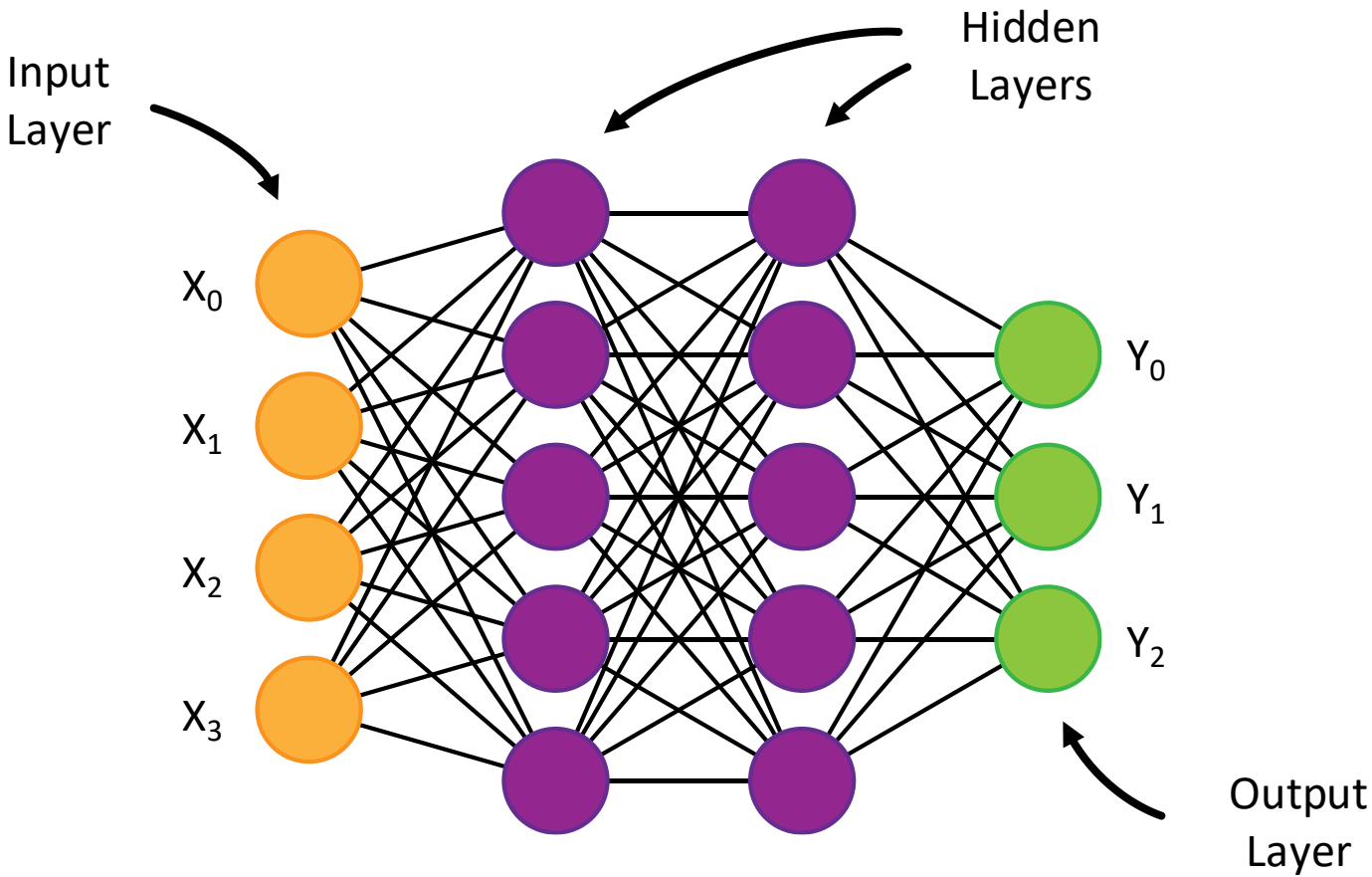
A dank meme is exceptionally odd or unique

Given the following context:

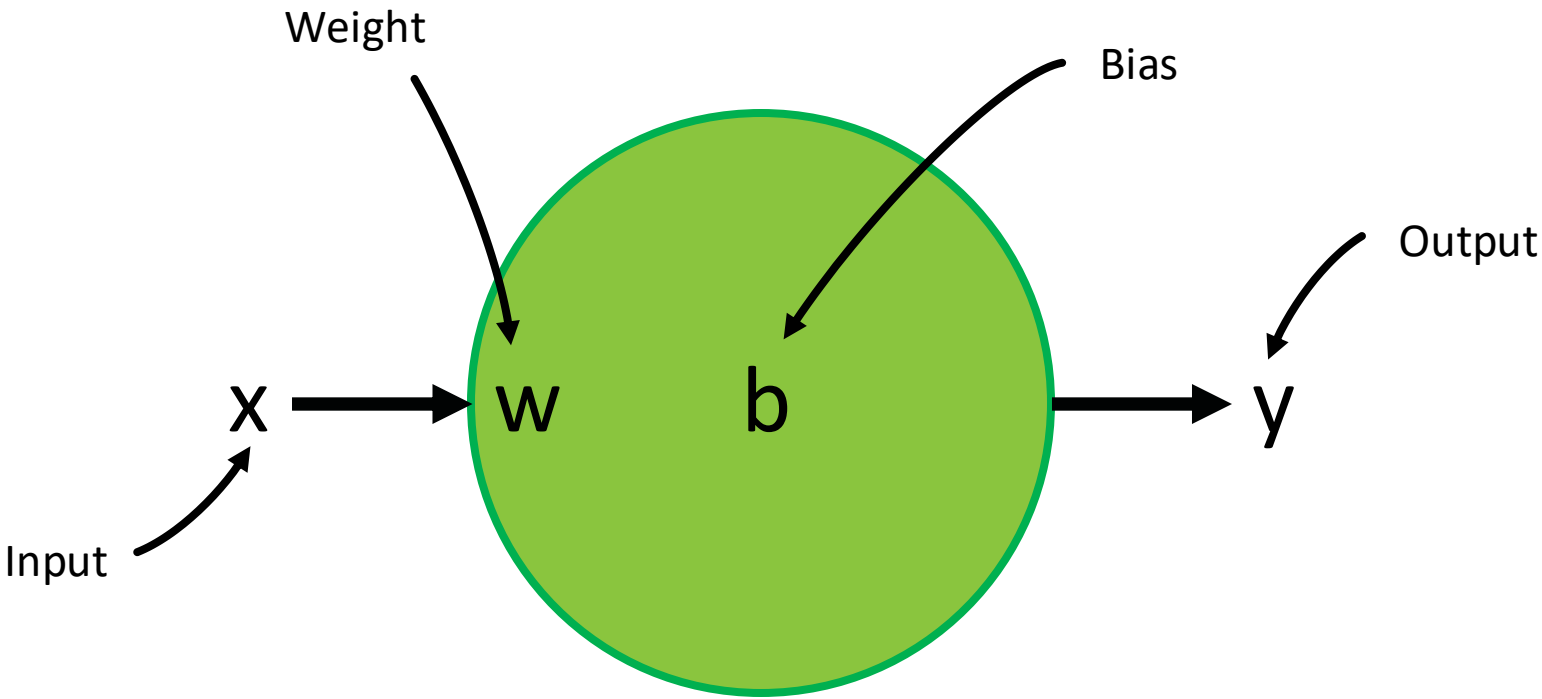
Answer the following question:

Can you explain to me what a dank meme is?

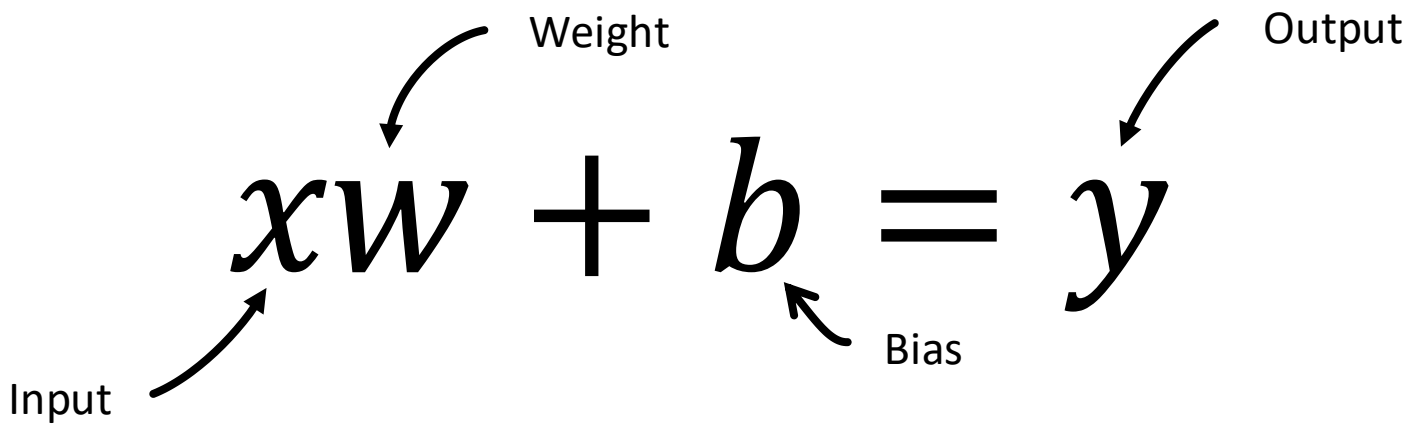

CREATING EMBEDDING MODELS YOURSELF
USING EXISTING EMBEDDING MODELS

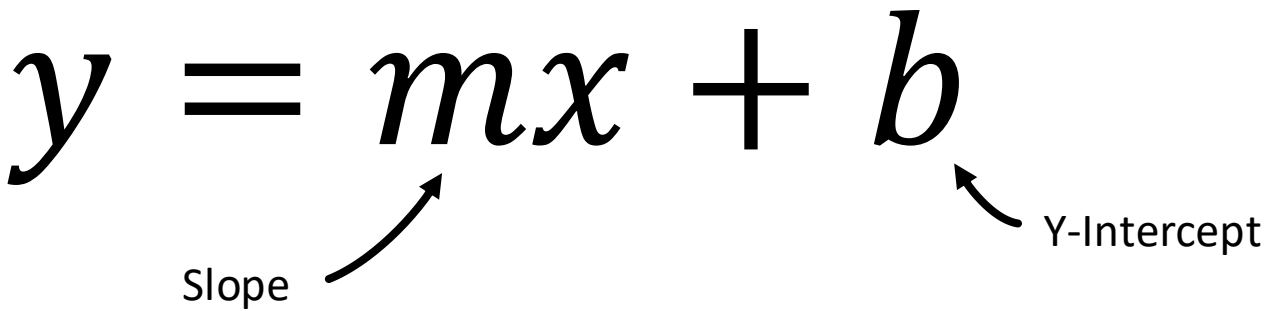


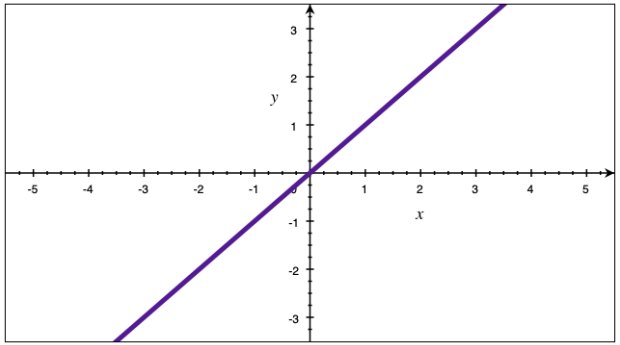
YOUR BASIC NEURAL NETWORK

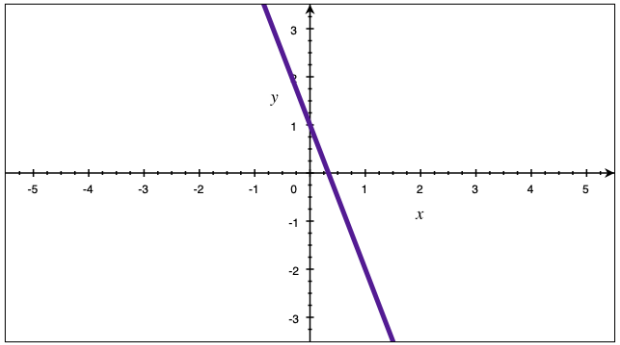


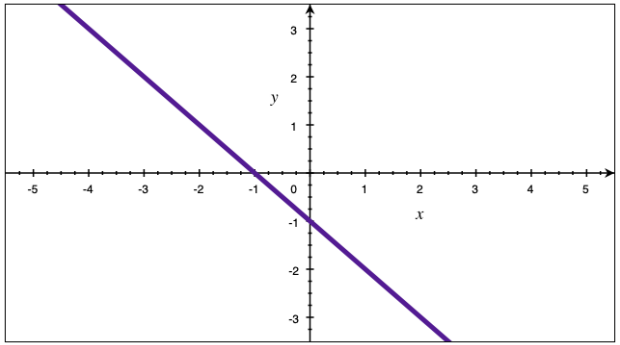
YOUR BASIC NEURON

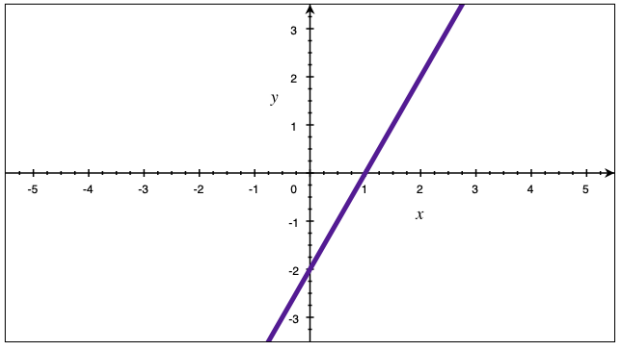

Weight
Output
$xw + b = y$
Input
Bias


$y = mx + b$
Slope
Y-Intercept

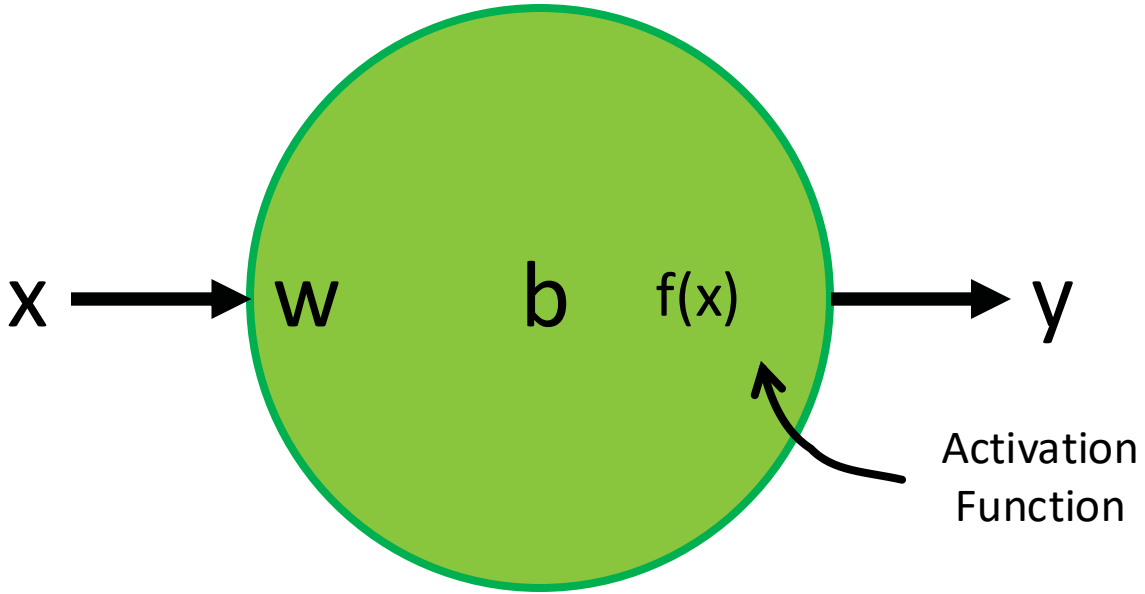




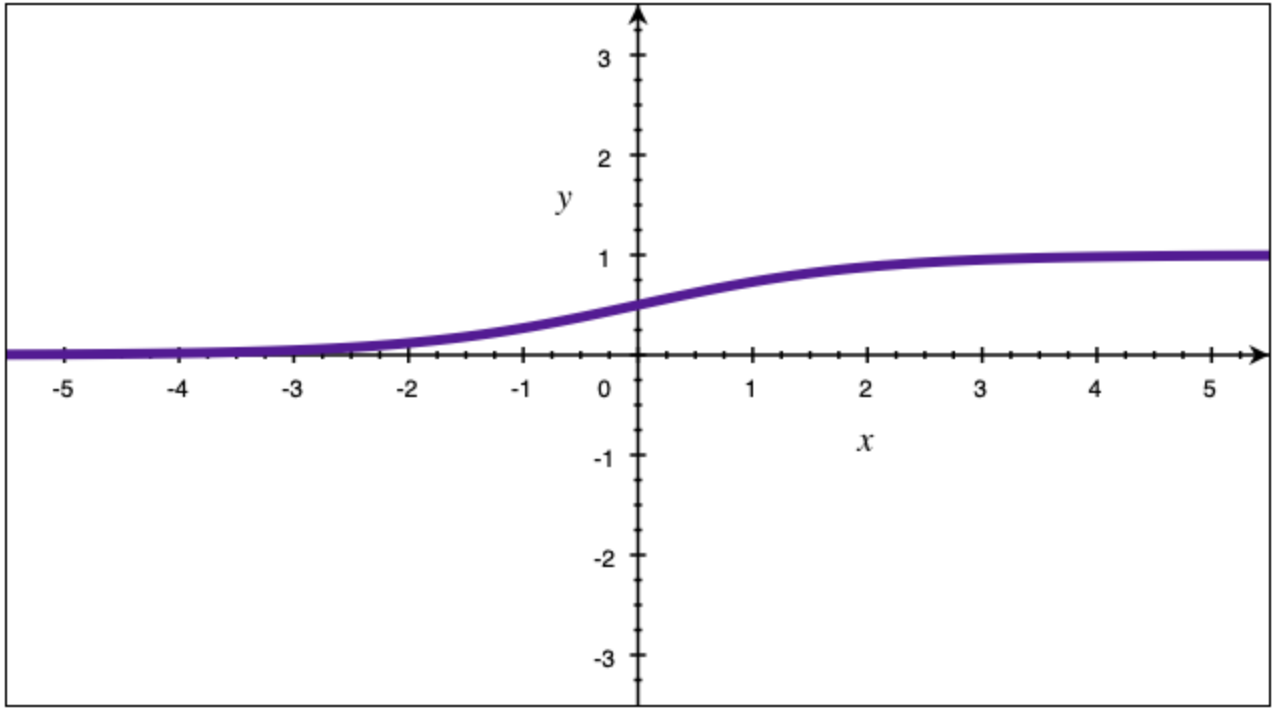




LINEAR FUNCTIONS

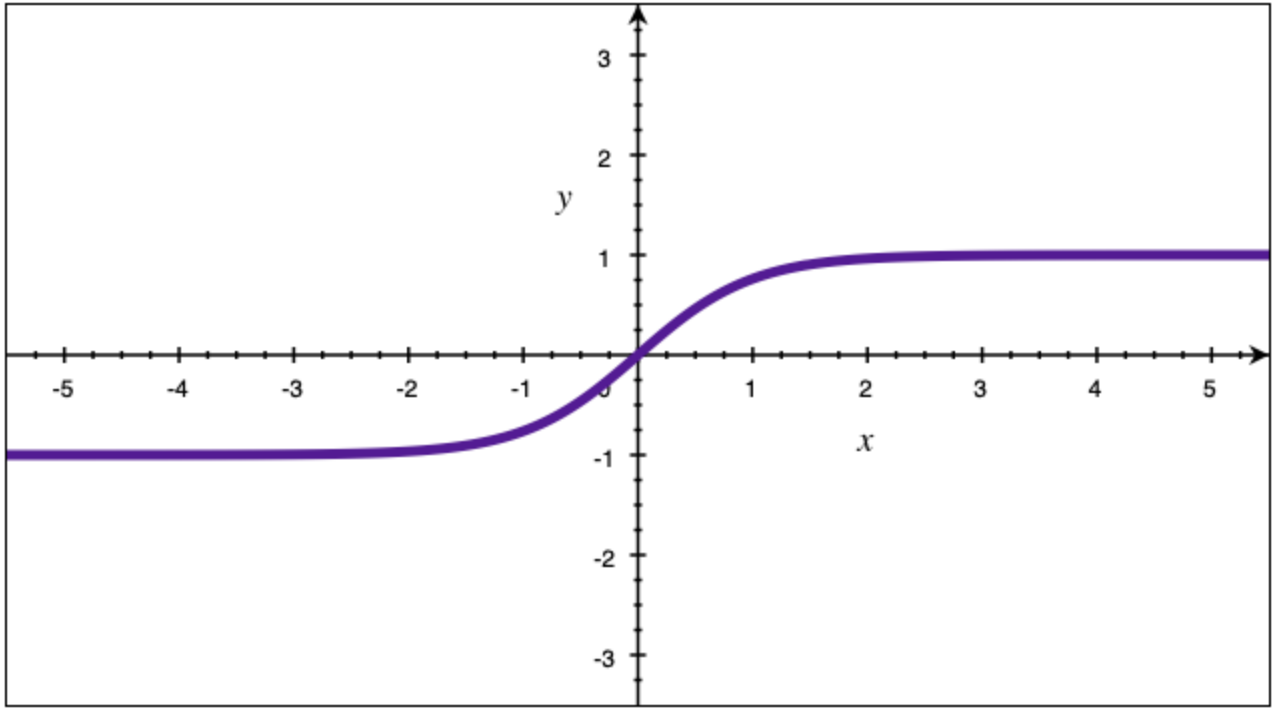


LESS LINEAR FUNCTIONS

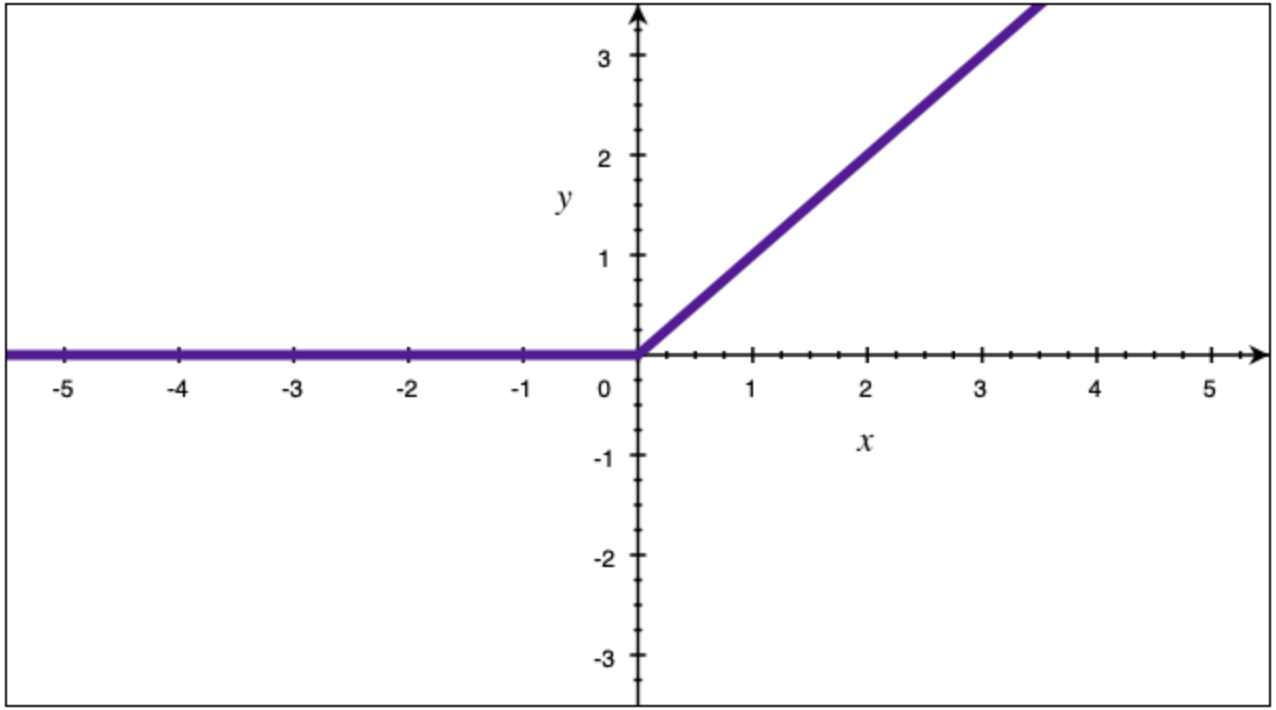


SIGMOID

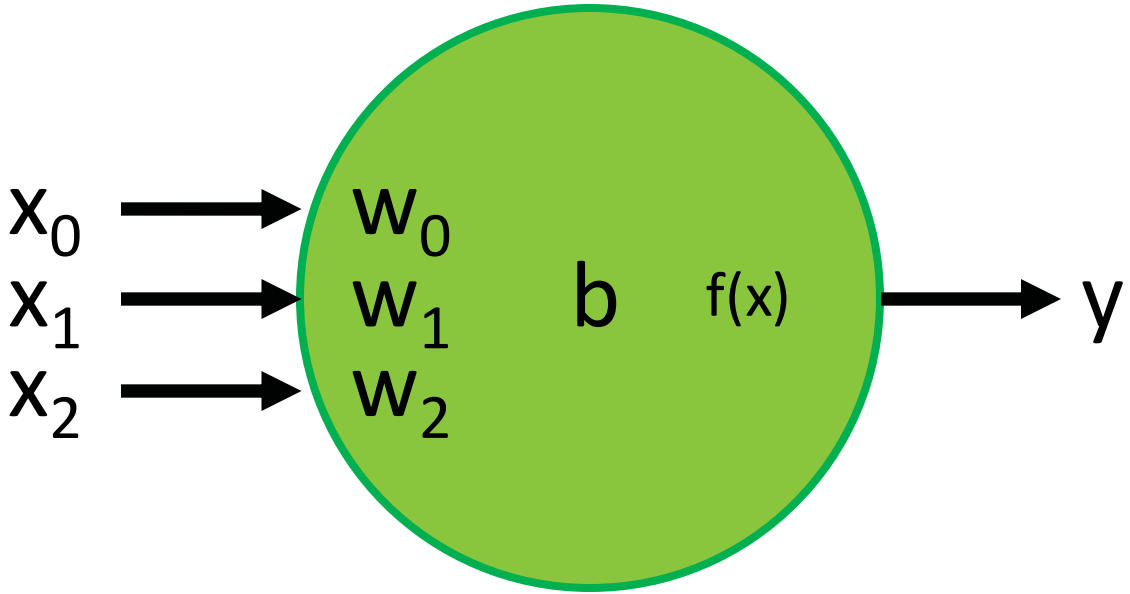
Redis



HYPERBOLIC TANGENT



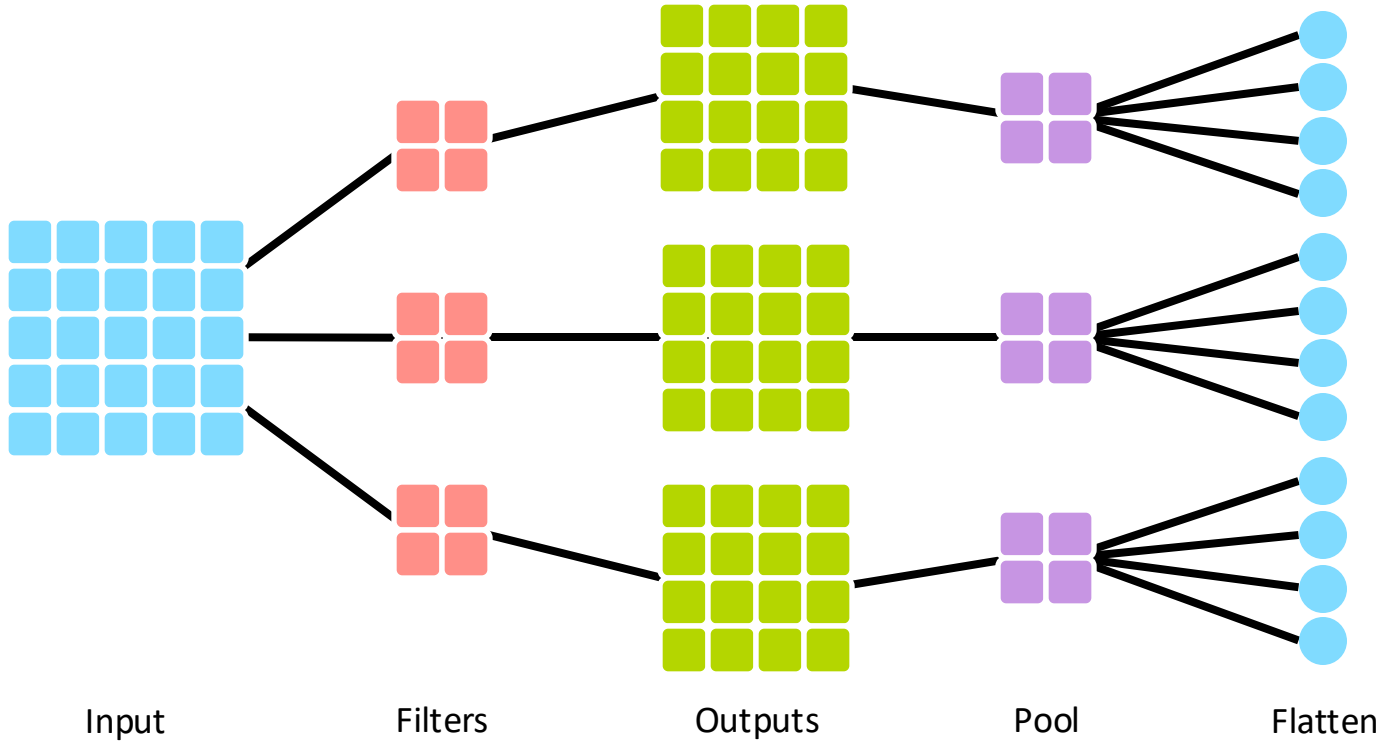
RECTIFIED LINEAR UNITS (ReLU)



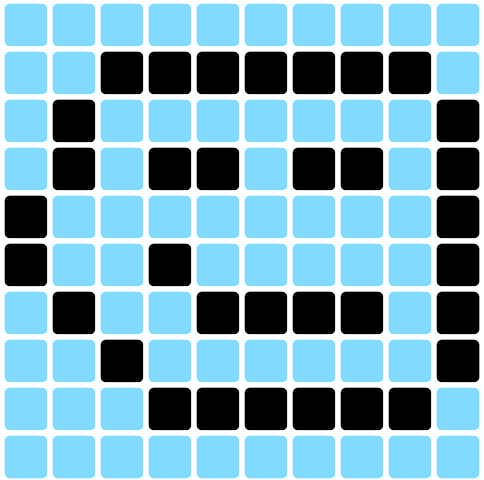
A MORE REALISTIC NEURON

$$x_0 w_0 + x_1 w_1 + x_2 w_2 + b = y$$

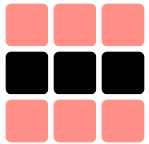
$$\sum_i x_i w_i + b$$



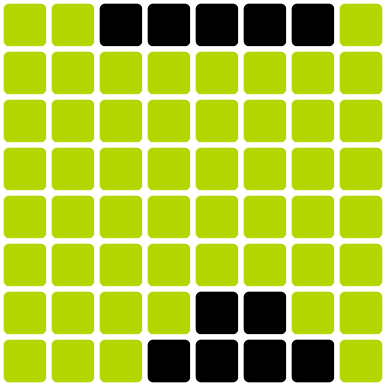
CONVOLUTION NEURAL NETWORK



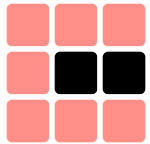
Input



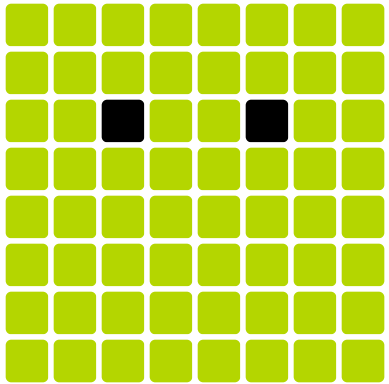
Filter



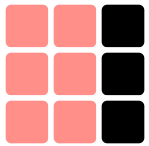
Output



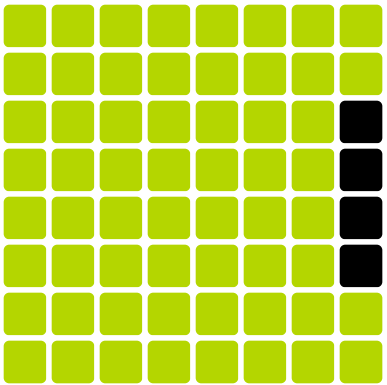
Filter



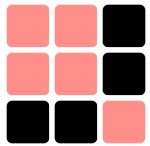
Output



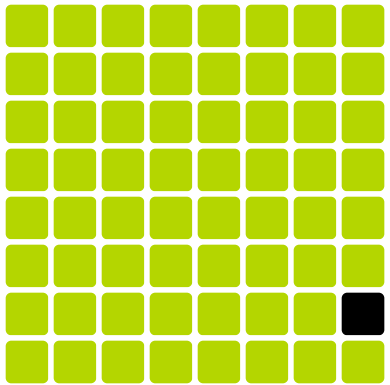
Filter



Output

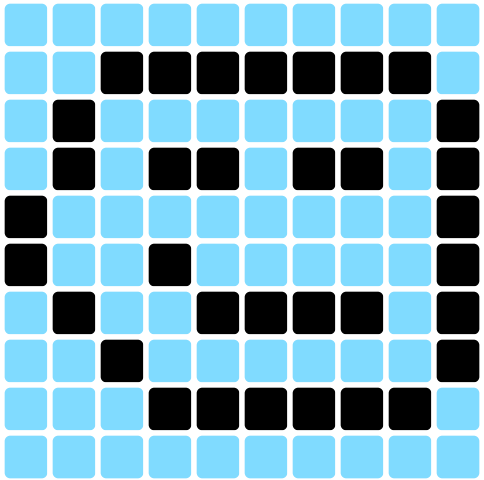


Filter

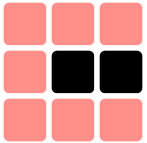


Output

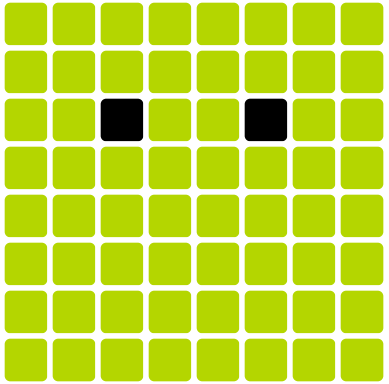
Redis



Input

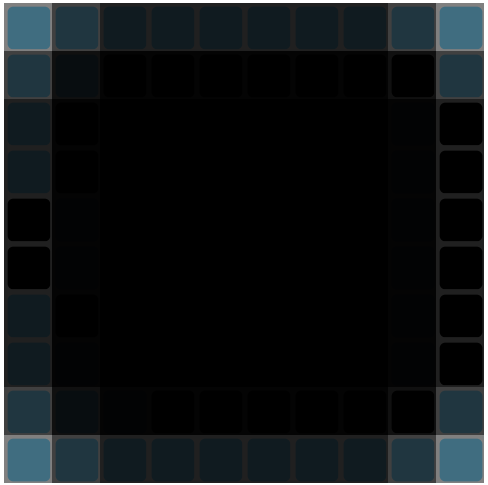


Filter

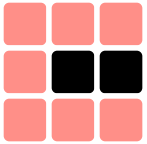


Output

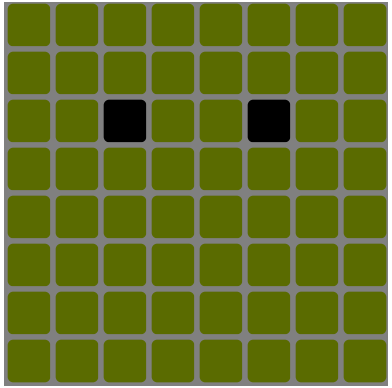
CONVOLUTIONS



Input

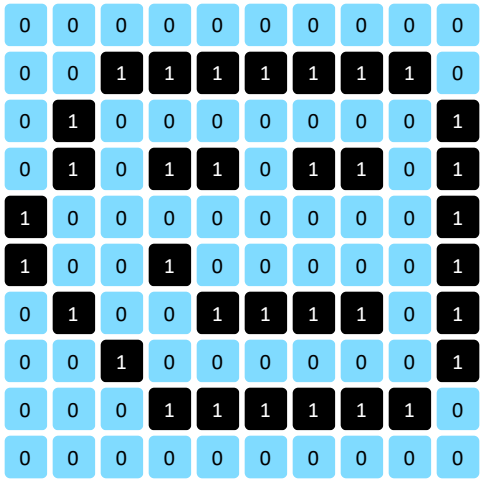


Filter



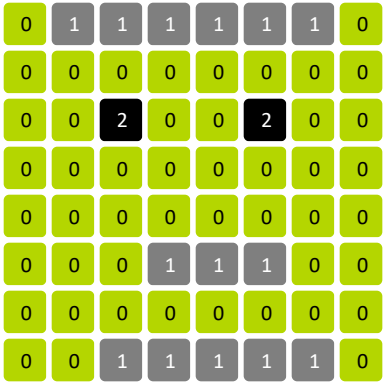
Output

CONVOLUTIONS



Input

Filter



Output

CONVOLUTIONS WITH MATH

Redis

x_i

0	0	0	0	0	0	0	0	0	0
0	0	1	1	1	1	1	1	1	0
0	1	0	0	0	0	0	0	0	1
0	1	0	1	1	0	1	1	0	1
1	0	0	0	0	0	0	0	0	1
1	0	0	1	0	0	0	0	0	1
0	1	0	0	1	1	1	1	0	1
0	0	1	0	0	0	0	0	0	1
0	0	0	1	1	1	1	1	1	0
0	0	0	0	0	0	0	0	0	0

w_i

-1	-1	-1
-1	1	1
-1	-1	-1

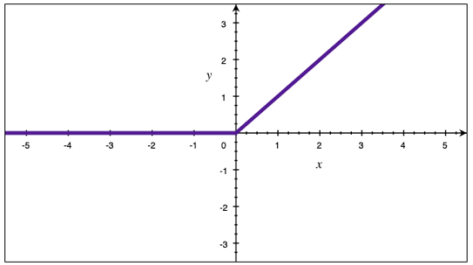
b

0

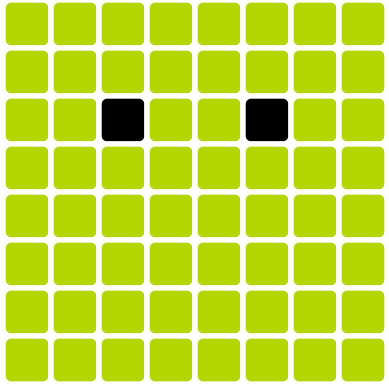
y_i

0	1	1	1	1	1	1	0
0	0	0	0	0	0	0	0
0	0	2	0	0	2	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	1	1	1	0	0
0	0	0	0	0	0	0	0
0	0	1	1	1	1	1	0

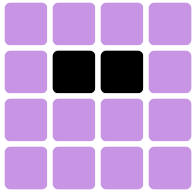
$$\sum_i x_i w_i + b$$



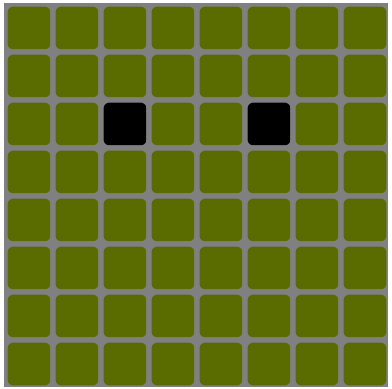
Redis



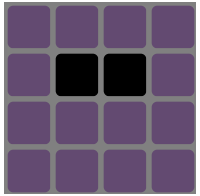
2x2



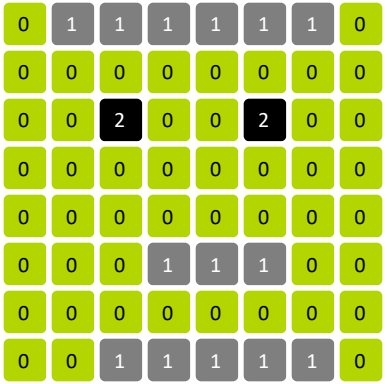
POOLING



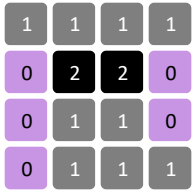
2x2



POOLING

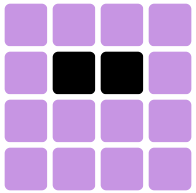


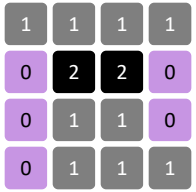
2x2



POOLING

Redis





FLATTENING

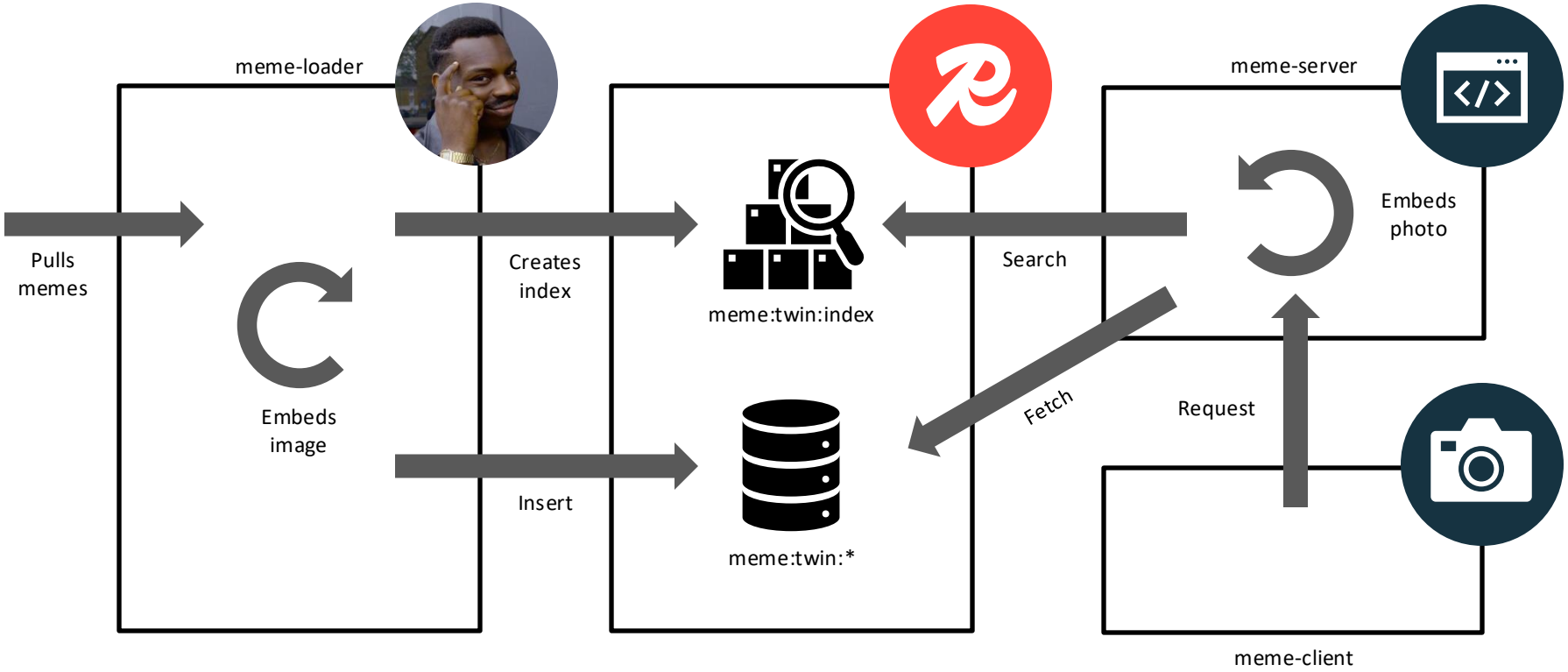
Redis


CUSTOM EMBEDDING MODELS
DEVS
EXISTING EMBEDDING MODELS
Redis


ONE DOES NOT SIMPLY
CREATE THEIR OWN EMBEDDING MODELS
Redis


Y'ALL GOT ANY MORE OF THOSE
VECTOR DATABASE DEMOS
Redis


I'M NOT SAYING I'M GONNA USE REDIS
BUT I'M GONNA USE REDIS
HISTORY.COM
Redis


meme-loader
Pulls memes
Embeds image
Creates index
Insert
meme:twin:index
meme:twin:*
Search
Fetch
meme-server
Embeds photo
Request
meme-client


ME WHEN I SEE A TALK BUT BEFORE THE DEMO

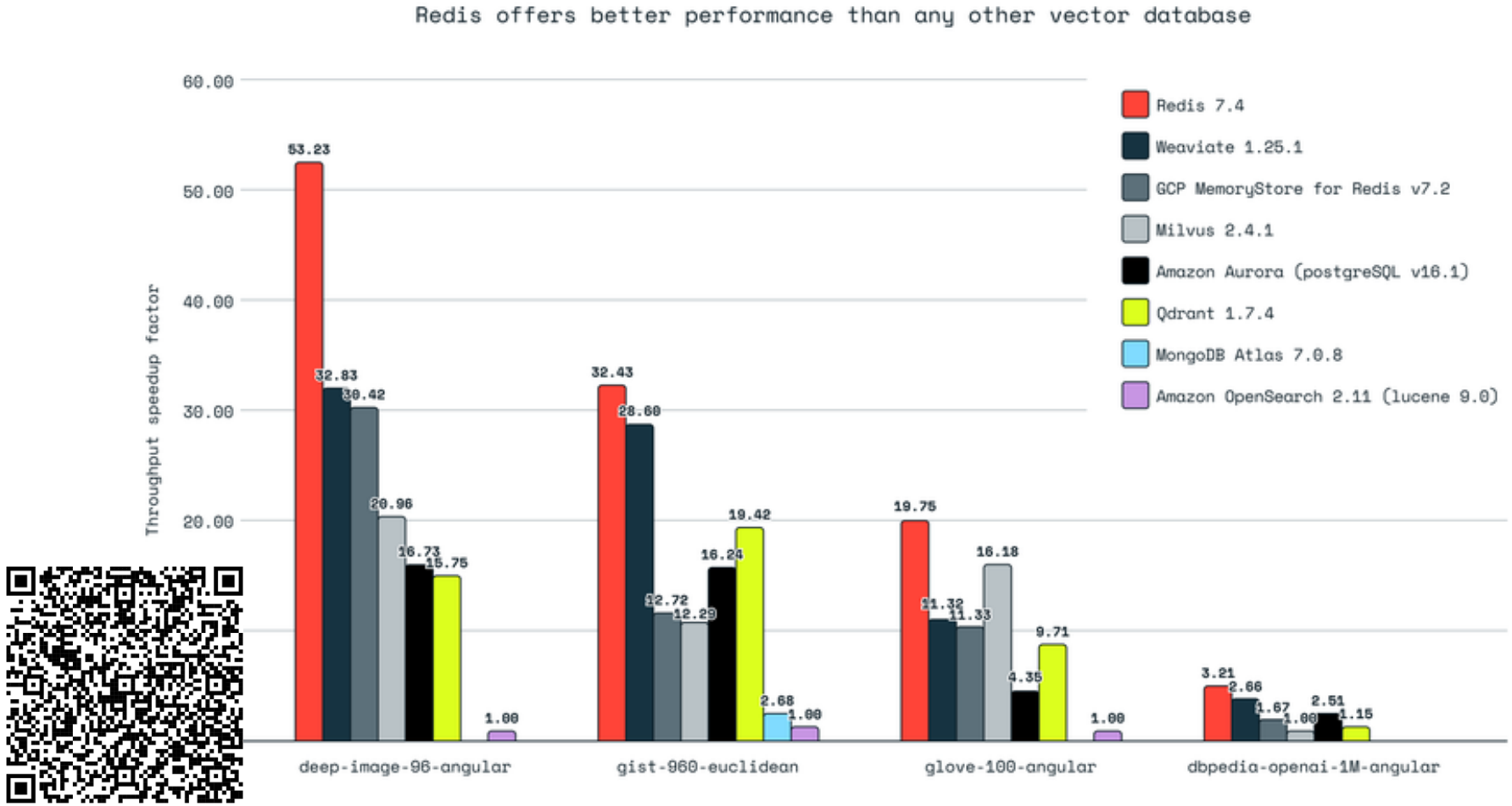
github.com/guyroyse/
modern-problems-require-modern-solutions

Redis

redis.io
university.redis.io

Discord

discord.gg/redis


Redis offers better performance than any other vector database
Throughput speedup factor
60.00
50.00
40.00
30.00
20.00
Redis 7.4
Weaviate 1.25.1
GCP MemoryStore for Redis v7.2
Milvus 2.4.1
Amazon Aurora (postgreSQL v16.1)
Qdrant 1.7.4
MongoDB Atlas 7.0.8
Amazon OpenSearch 2.11 (lucene 9.0)
53.23
32.83
30.42
20.96
16.73
15.75
1.00
32.43
28.60
12.72
12.29
16.24
19.42
2.68
1.00
19.75
11.32
11.33
16.18
4.35
9.71
1.00
3.21
2.66
1.67
1.00
2.51
1.15
deep-image-96-angular
gist-960-euclidean
glove-100-angular
dbpedia-openai-1M-angular

Guy Royse

Developer Advocate

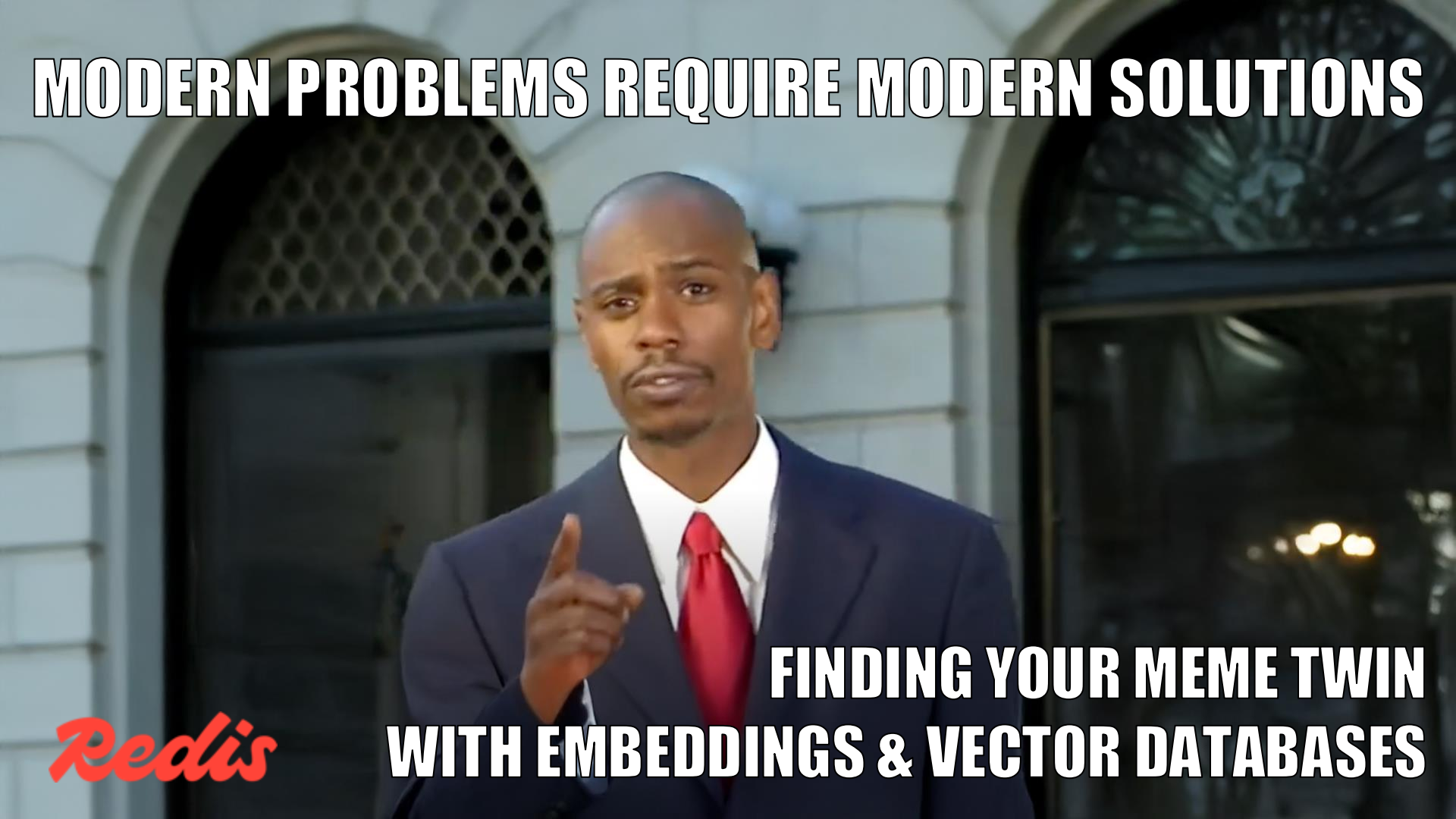
Redis

@guy.dev

@guyroyse

github.com/guyroyse

guy.dev


MODERN PROBLEMS REQUIRE MODERN SOLUTIONS
FINDING YOUR MEME TWIN
WITH EMBEDDINGS & VECTOR DATABASES
Redis