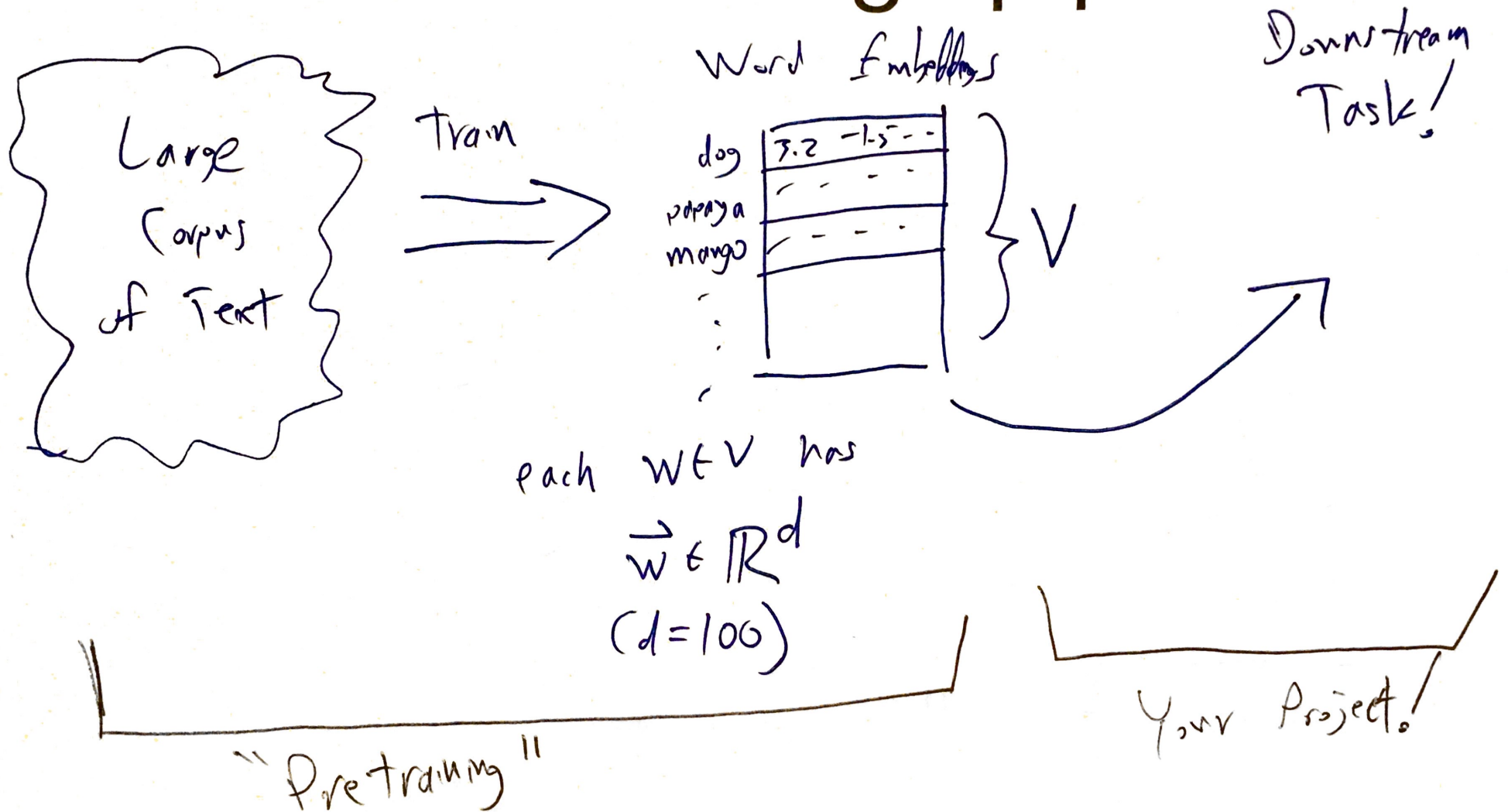


# Word embeddings pipeline



- Word embeddings are a lexical resource, to be used for downstream tasks
  - Transfer learning: get info huge corpus, then apply to learn from a small labeled dataset
- Compare to lexicons, lexical knowledge bases like WordNet, etc.

$x$  vectors

$y$

---

$$x^T y = \sum_{j=1}^d x_j y_j$$

$\text{cos sim}(x, y)$

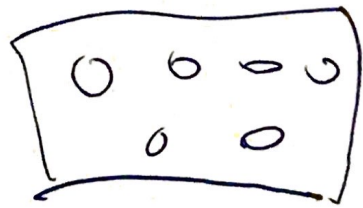
"Normalized inner product"

$$= \frac{x^T y}{\|x\| \|y\|}$$

$$\|x\| = \sqrt{\sum_j x_j^2}$$

# Application: document embedding

- Instead of bag-of-words, can we derive a latent embedding of a document/sentence?



$$V_i \in \mathbb{R}^d$$

$$\text{"Bag of Embeddings"} = \frac{1}{N_{\text{tok}}} \sum_{i=1}^{N_{\text{tok}}} V_i = X$$

(classify) using  $X$  as the feat. vector

$$P(y=1/x) = \frac{1}{1 + e^{-\beta^T x}}$$

$$\text{BoE: } X \in \mathbb{R}^{100}$$

$$(d=100)$$

$$\text{BoW: } X \in \mathbb{R}^V$$

$$(V=20,000)$$

(Log Reg)

Pros of BoE

- Quicker

- Generalize to quasi-synonyms

Con of BoE

- Embeddings are bad.

Antonym!

- D. main momatch