CS21SI: AI for Social Good

Lecture 1: Motivations and Basic Models

Instructors

Karan Singhal

Swetha Revanur

Shubhang Desai

Course Sponsor

Chris Piech

Course Staff

Daniel Wu

Email us at cs21si-staff@lists.stanford.edu!

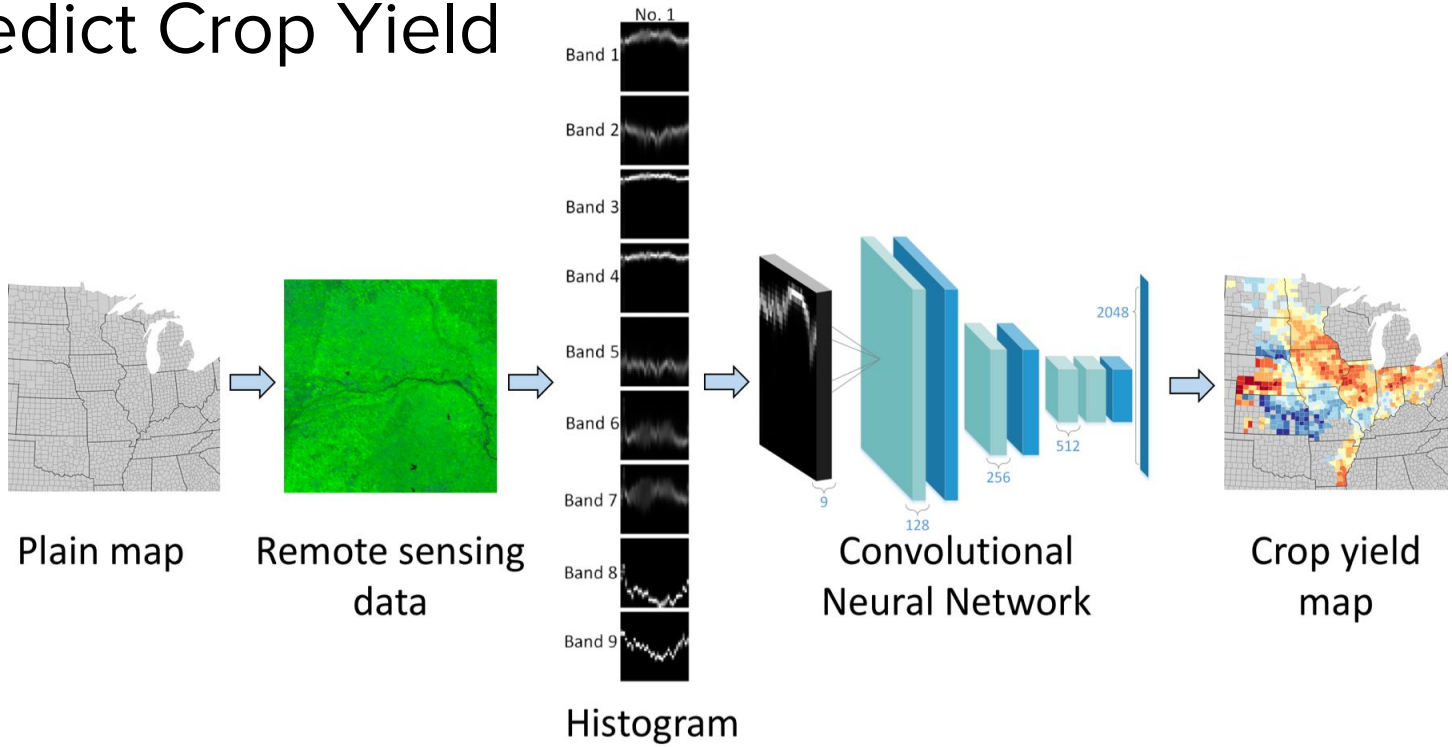
Why AI for Social Good?

Why machine learning & deep learning?

- Neural networks can simply model complex data
- More computational power
- More data
- Better models and training techniques
- Low hanging fruit!

Predicting Poverty from Satellite Images

Combining Remote Sensing Data and Machine Learning to Predict Crop Yield



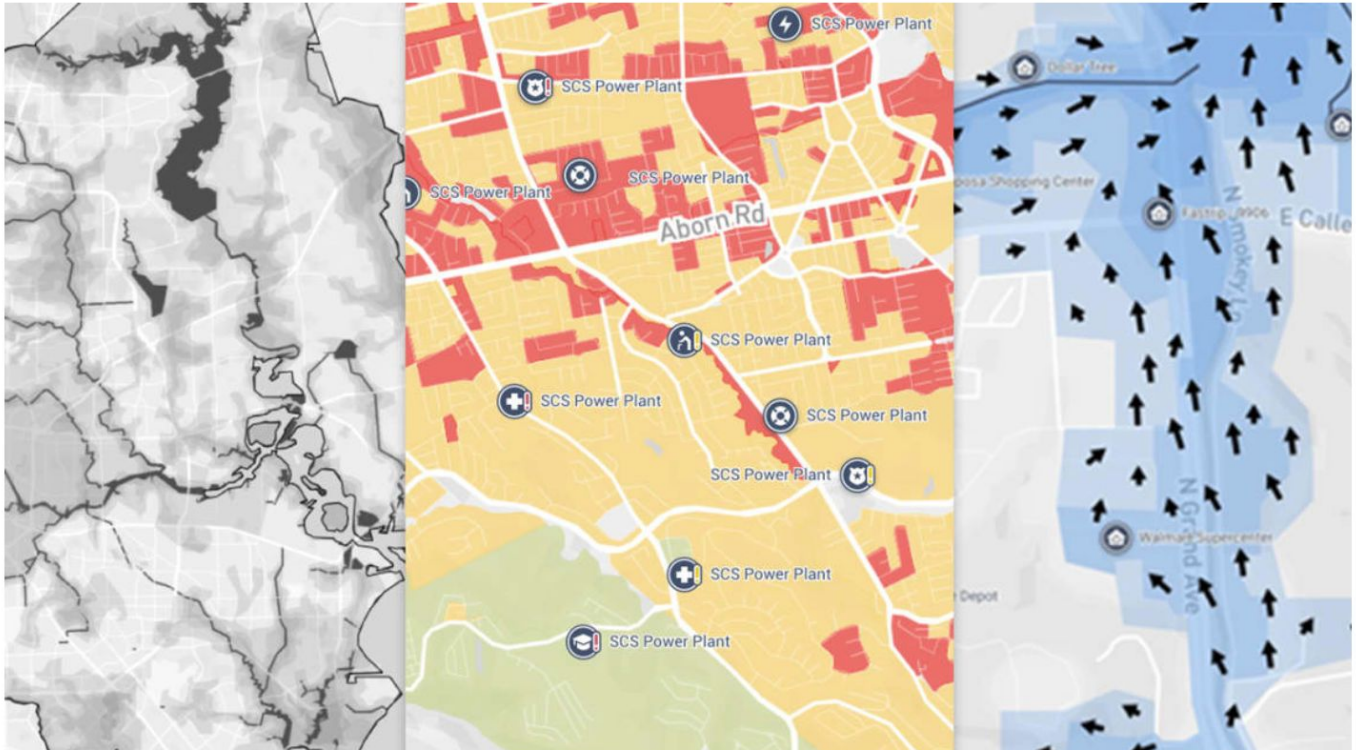
How data scientists are using AI for suicide prevention

The Crisis Text Line uses machine learning to figure out who's at risk and when to intervene.

By Brian Resnick | @B_resnick | brian@vox.com | Updated Jun 9, 2018, 7:22am EDT

Disaster relief is dangerously broken. Can AI fix it?

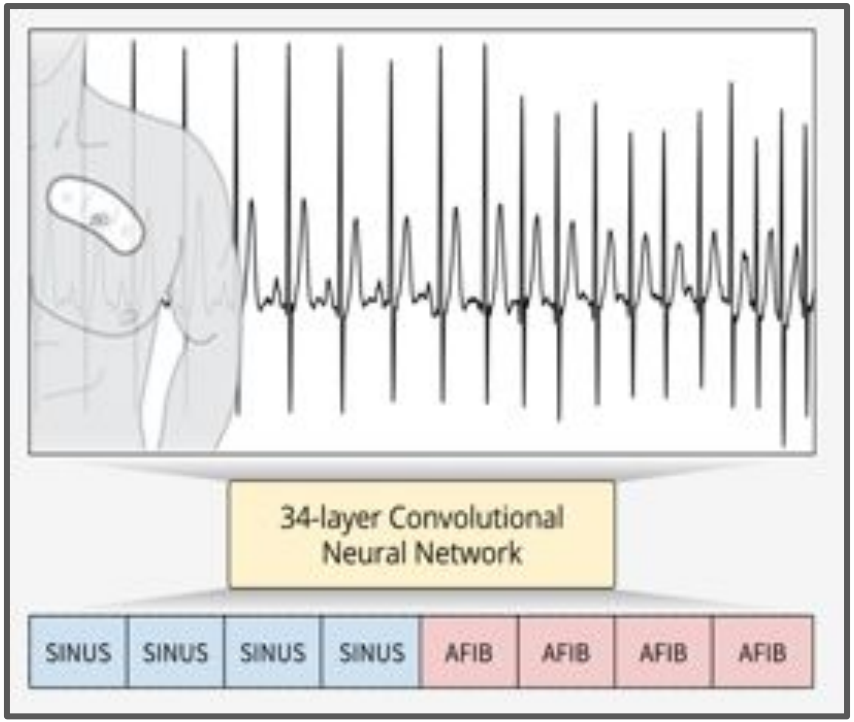
Cities are looking to machine learning to streamline their disaster-response efforts. Will it be too little too late?



Deep Knowledge Tracing

Chris Piech[*], **Jonathan Bassen**[*], **Jonathan Huang**[*‡], **Surya Ganguli**[*],
Mehran Sahami[*], **Leonidas Guibas**[*], **Jascha Sohl-Dickstein**[*†]
[*]Stanford University, [†]Khan Academy, [‡]Google
{piech,jbassen}@cs.stanford.edu, jascha@stanford.edu,

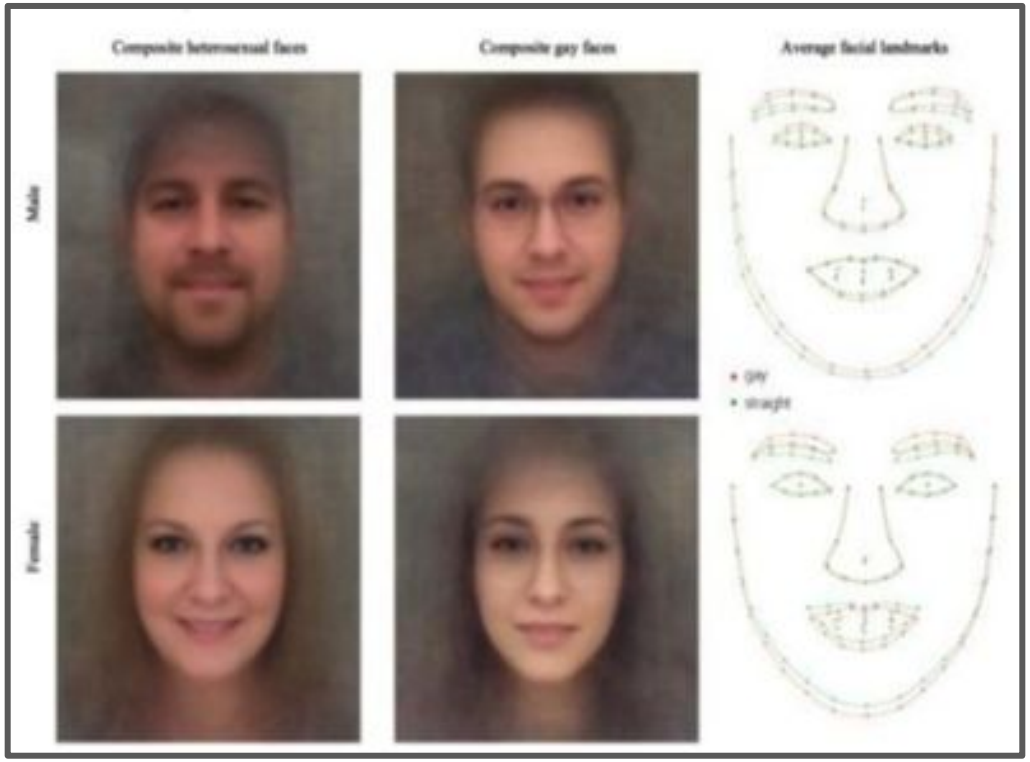
Using CNNs to Detect Arrhythmias



Problems with AI

- Biased data and models lead to biased predictions
- Engineers push cutting-edge, but not socially relevant bottom-line
- Automation could replace human jobs and further widen wealth disparity
- The power to impersonate people!
- Longer term risks associated with artificial general intelligence

Predicting Sexuality from Images



Self-Driving Cars

Better Language Models and their Implications (OpenAI)

SYSTEM PROMPT (HUMAN-WRITTEN)

In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.

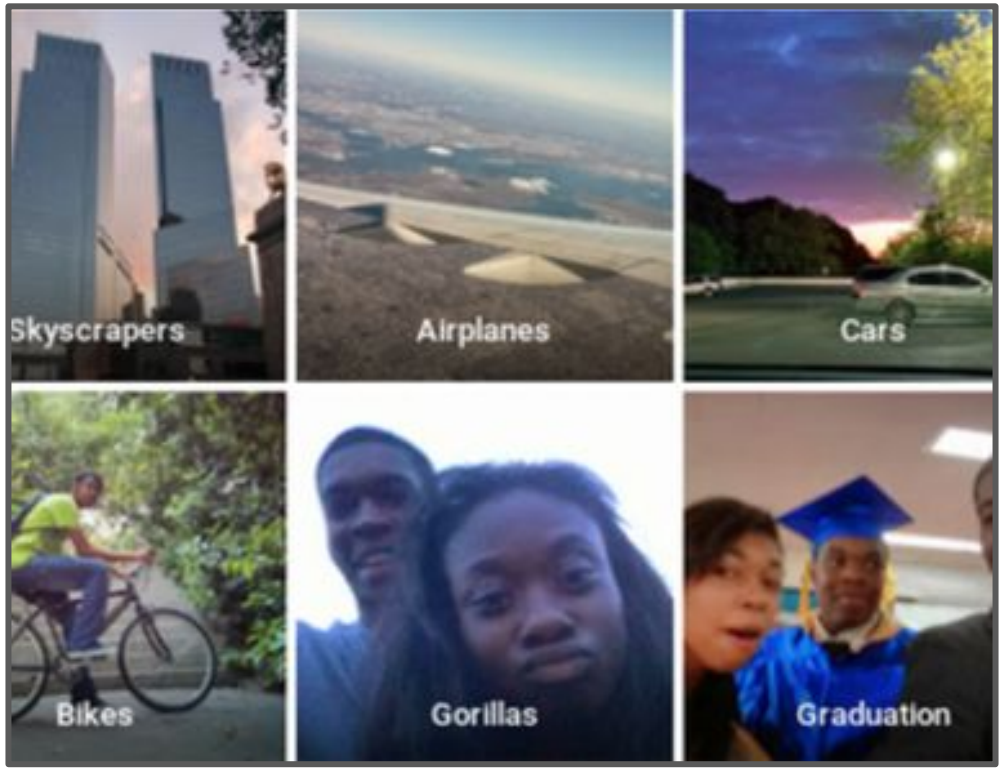
MODEL COMPLETION (MACHINE-WRITTEN, 10 TRIES)

The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

Labeling Pictures



Model Cards for Model Reporting

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, Timnit Gebru

(Submitted on 5 Oct 2018 (v1), last revised 14 Jan 2019 (this version, v2))

Class Goals

How to do AI

- Learn the techniques and some theory behind ML/DL
- Lectures and in-class exercises

How to do good with AI

- See and implement examples of AI being applied for good

How to not do bad with AI

- See examples of AI systems with negative implications

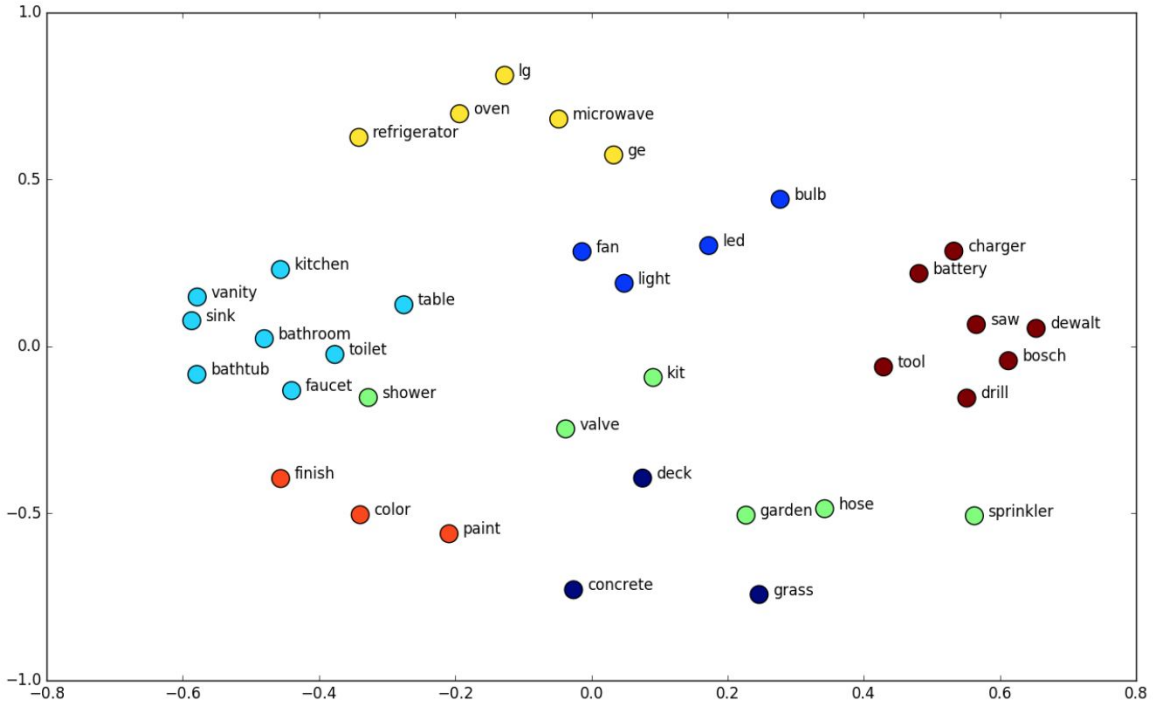
Logistics

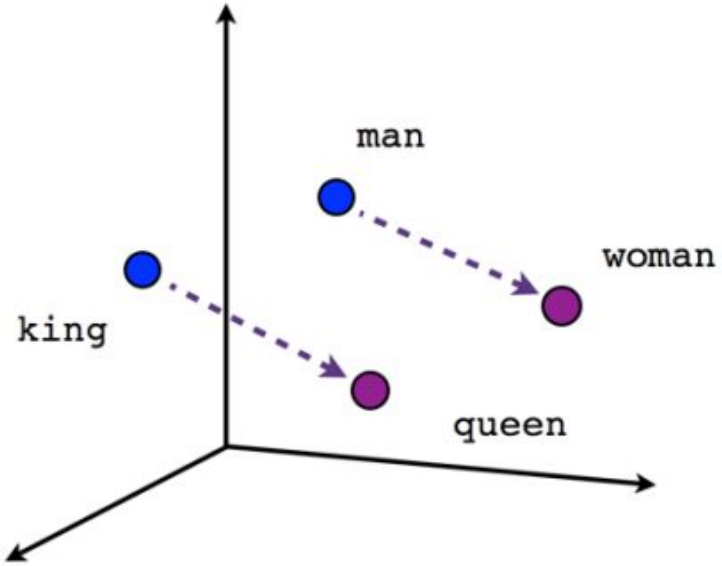
- Odd weeks are lectures from us, even weeks are speakers
  - Lectures will have in-class exercises!
- You are expected to be at all 10 class sessions (barring extreme circumstances)
- Weekly homework assignments in the form of iPython notebooks
  - Turn in homework as a PDF of CoLab notebook, emailed to CS 21si staff list
  - **Due 11:59pm on Monday (day before class)**
- 2 units, C/NC, meet once a week in this room (attendance required!)
- Class enrollment codes will be handed out Week 2

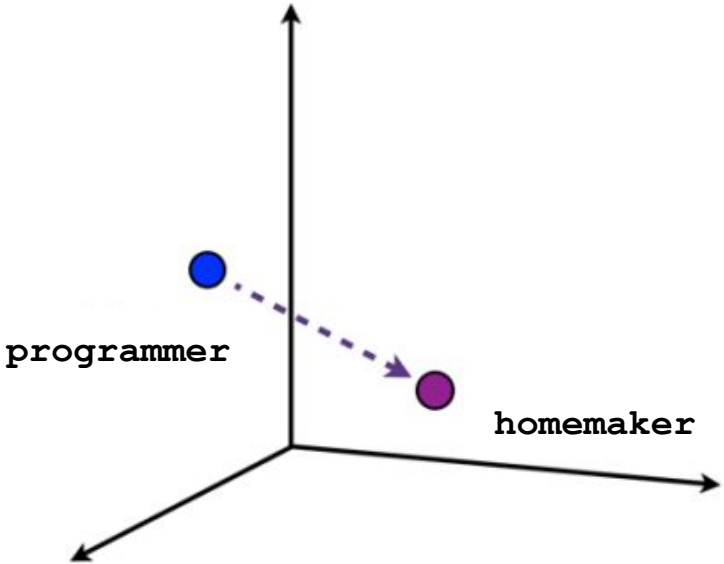
More Class Logistics

- Class website: cs21si.stanford.edu
- Class GitHub: github.com/karan1149/cs21si
- Contact us: cs21si-staff@lists.stanford.edu

Word Vectors




man
woman
king
queen


programmer
homemaker

Why do we care about word vectors?


Google


NETFLIX


Spotify®

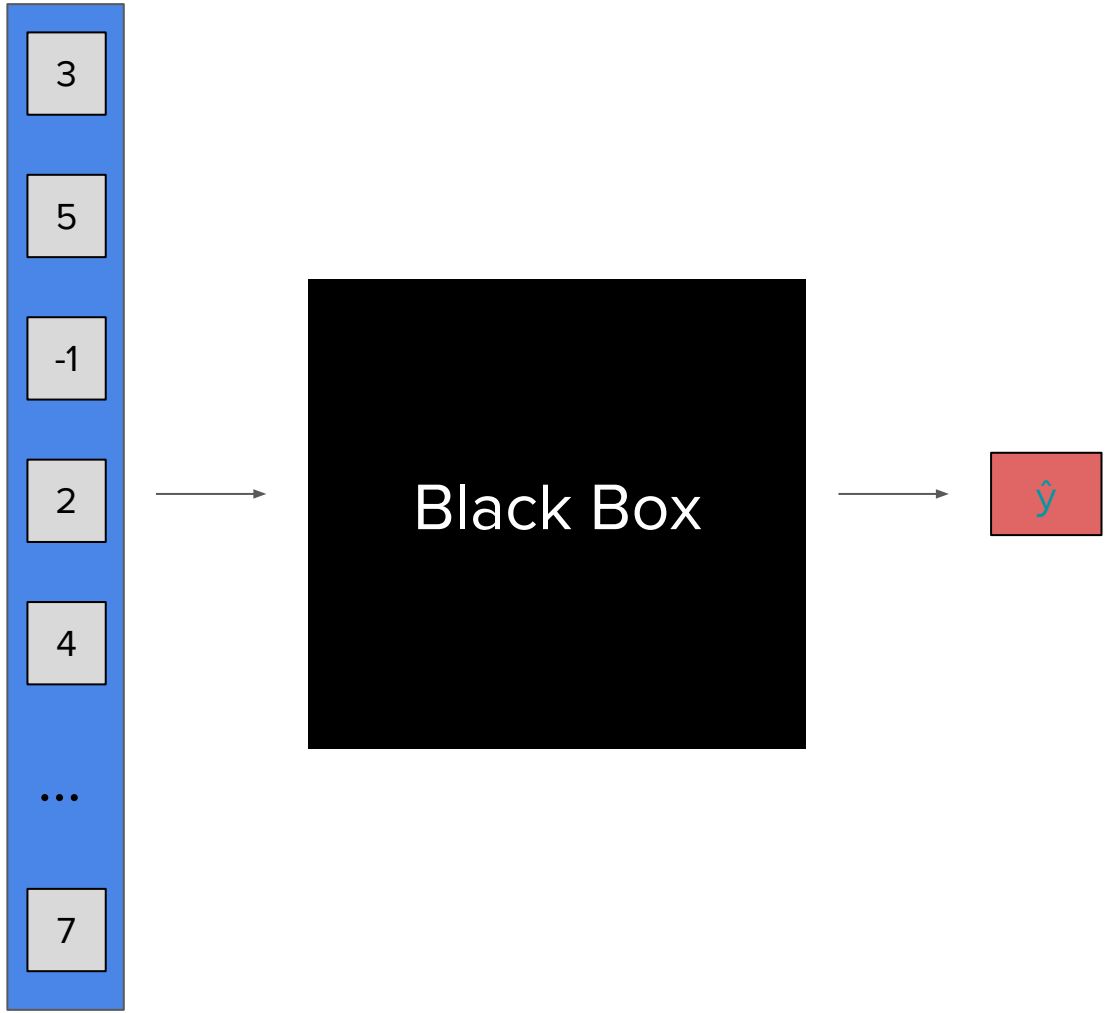

Google
Translate


YouTube

How bad is the problem?

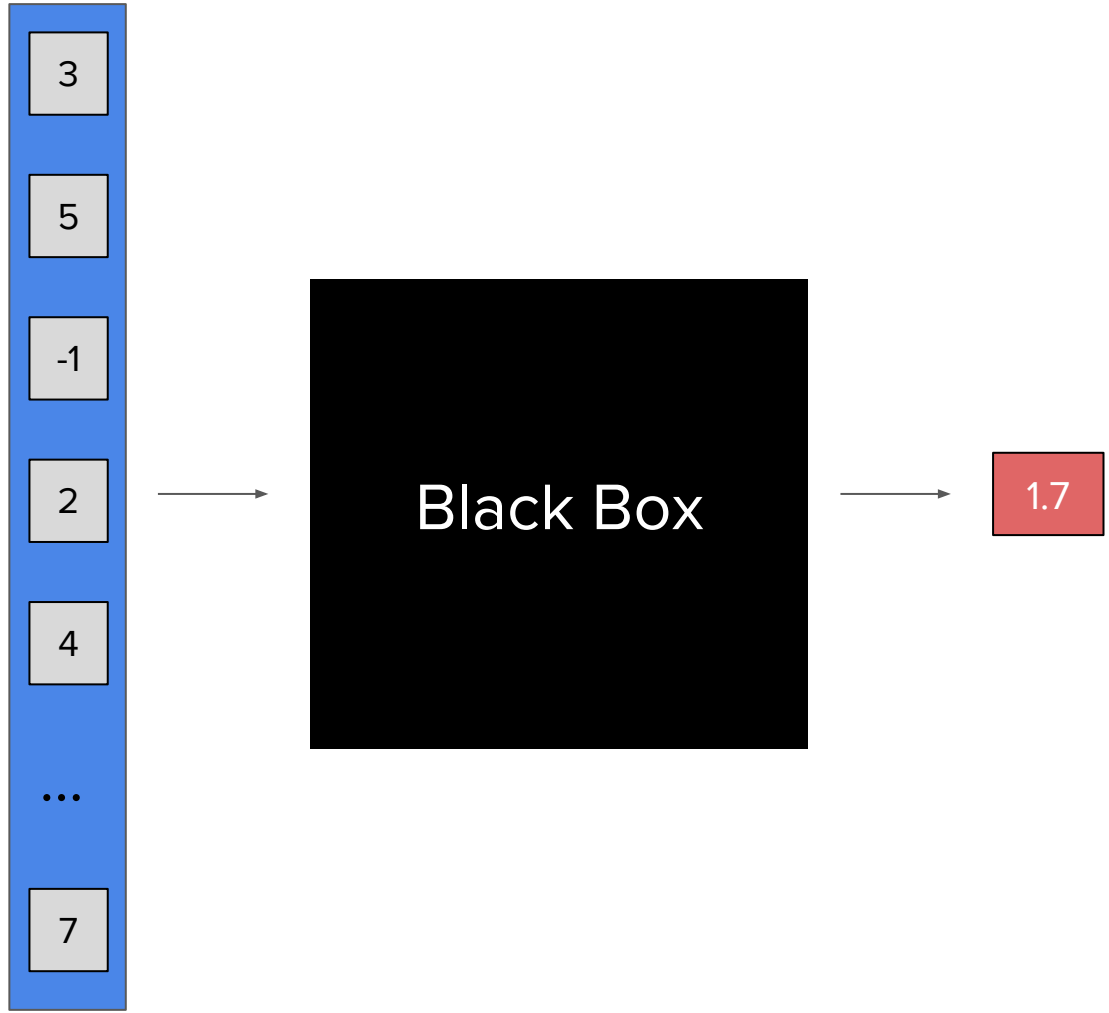
300-dimensional word vector

e.g. $v_{homemaker}$



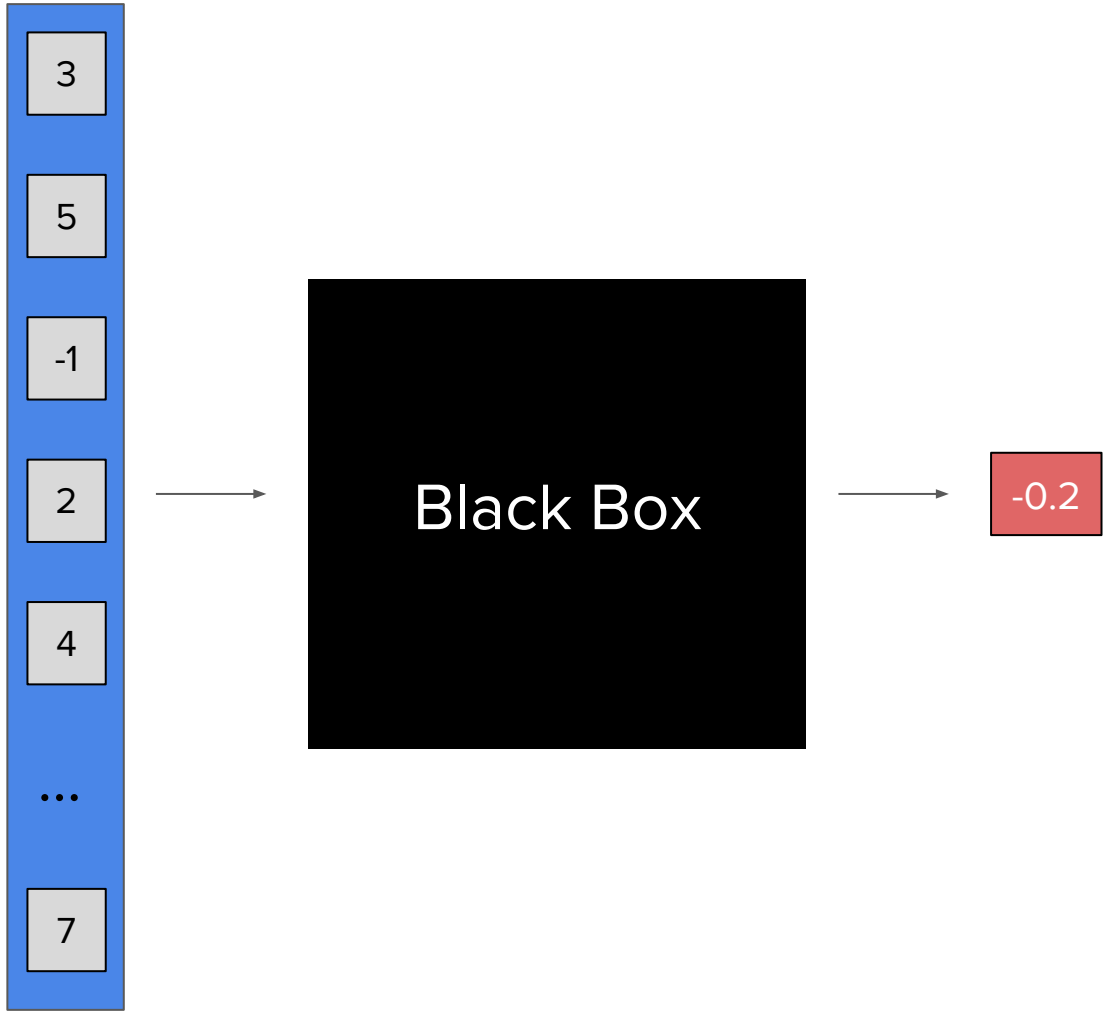
300-dimensional word vector

e.g. $v_{homemaker}$



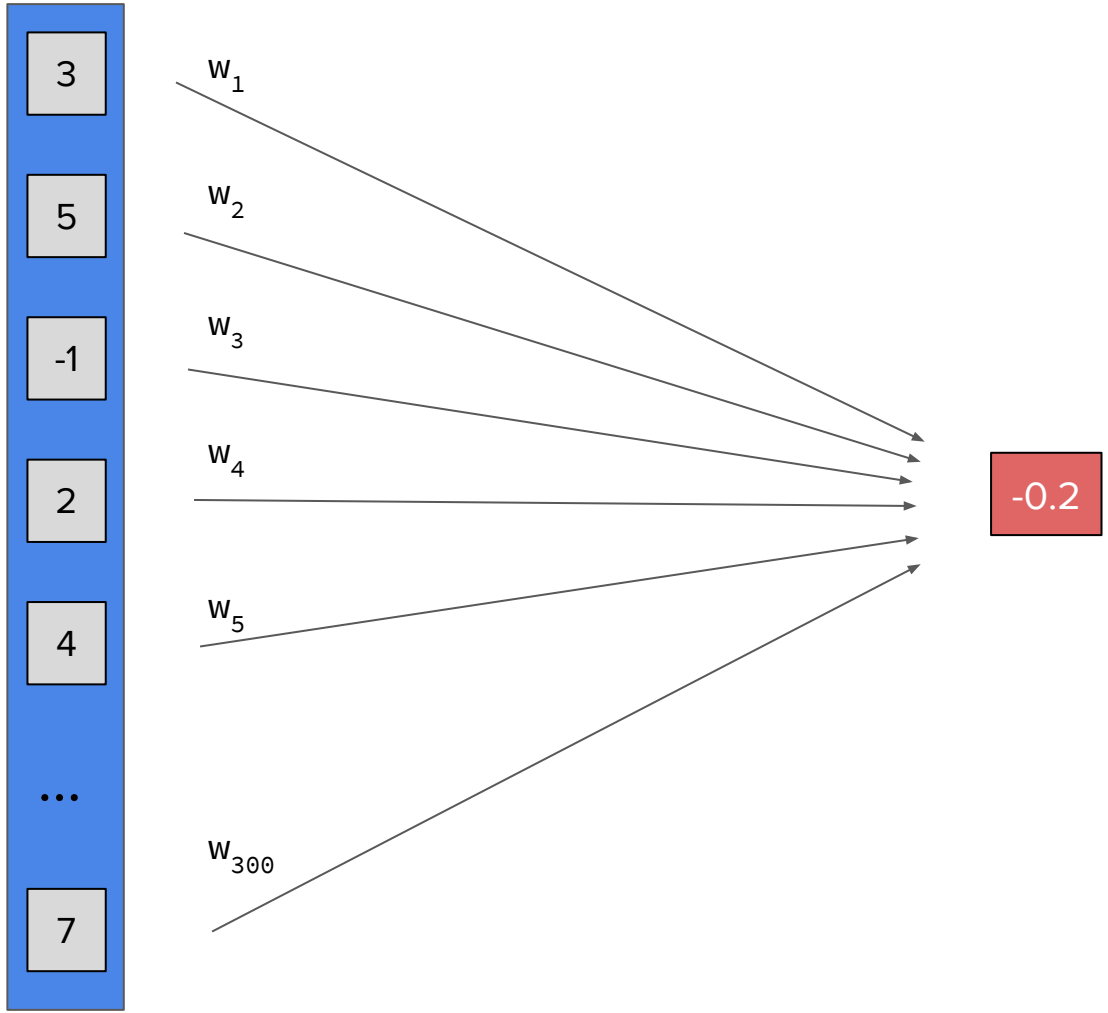
300-dimensional word vector

e.g. $v_{programmer}$



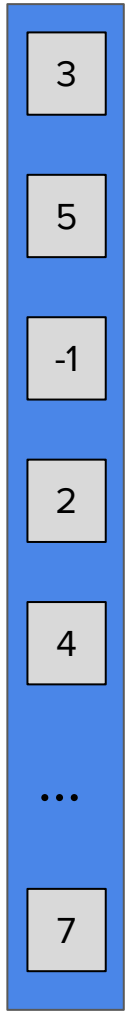
300-dimensional word vector

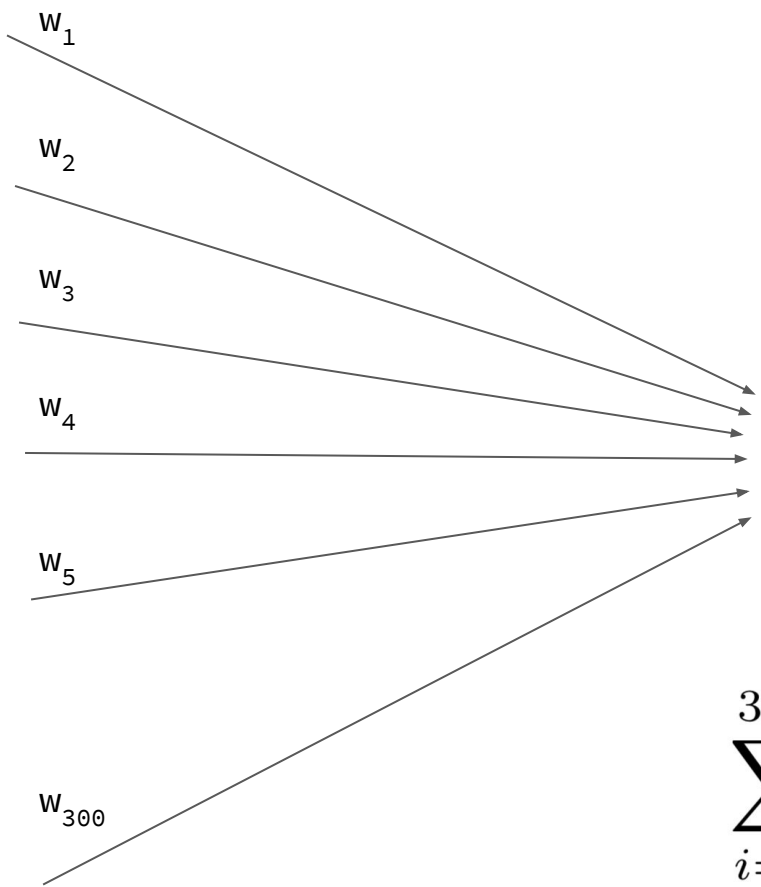
e.g. $v_{programmer}$



300-dimensional word vector

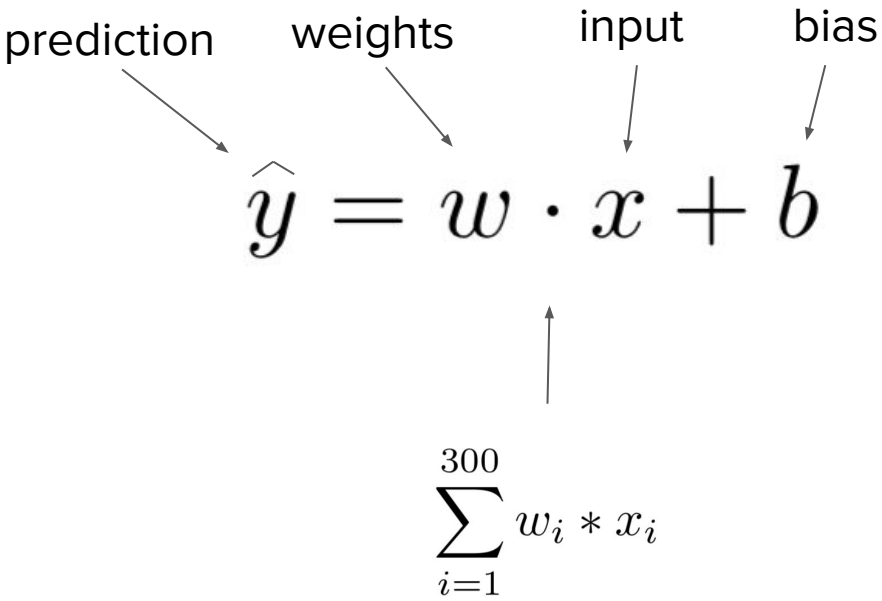
e.g. $v_{programmer}$





$$\sum_{i=1}^{300} w_i * x_i$$

Linear Model



How do we choose *w* and *b*?

$$\hat{y} = w \cdot x$$

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|}$$

$$\hat{y} = w \cdot x$$

This is a measure of the similarity between vectors *w* and *x*.

$$\hat{y} = w \cdot x$$

Choose $w = v_{woman} - v_{man}$!

Jupyter Notebook Exercises: Part 1

You'll need:

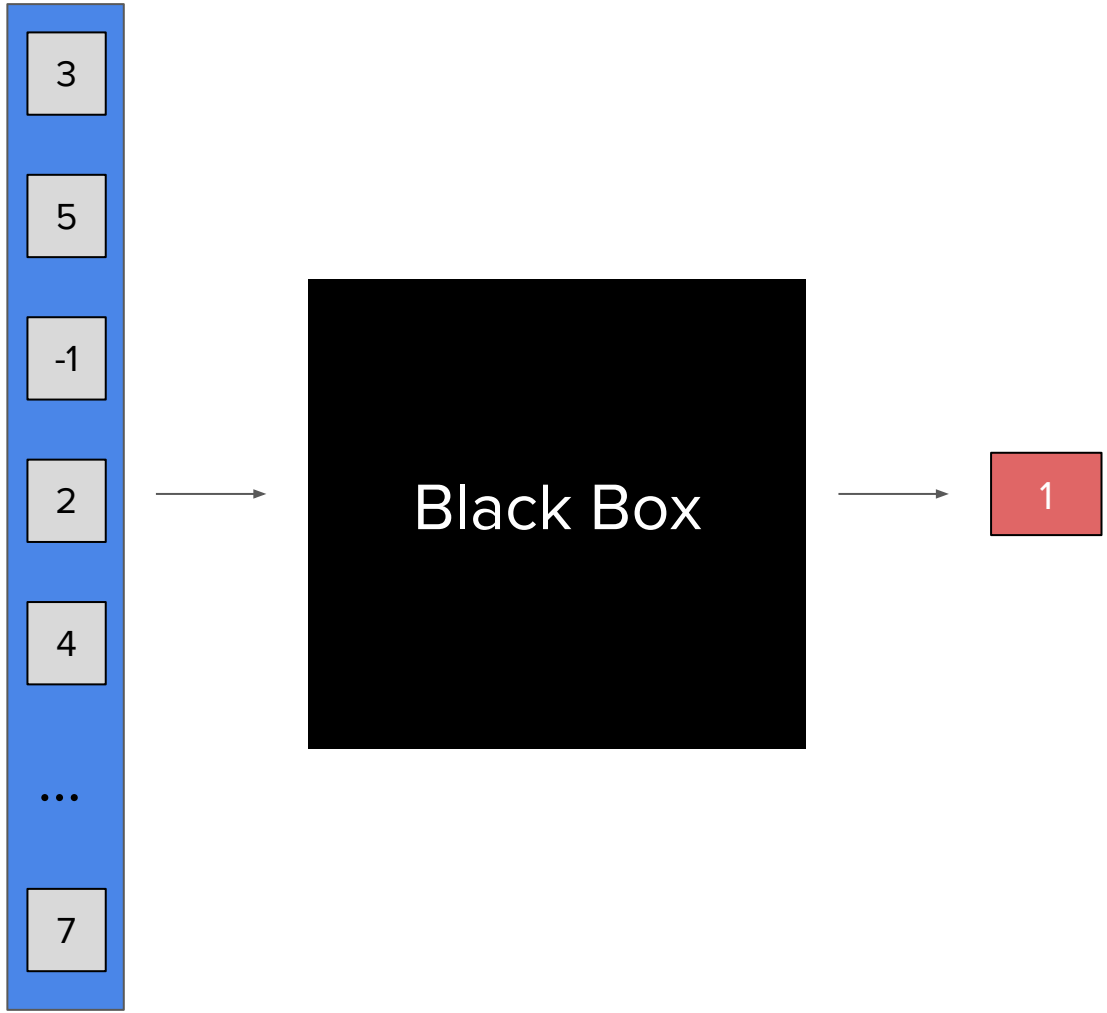
```
np.dot(a, b)

np.linalg.norm(a)
```

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

Improving upon our model

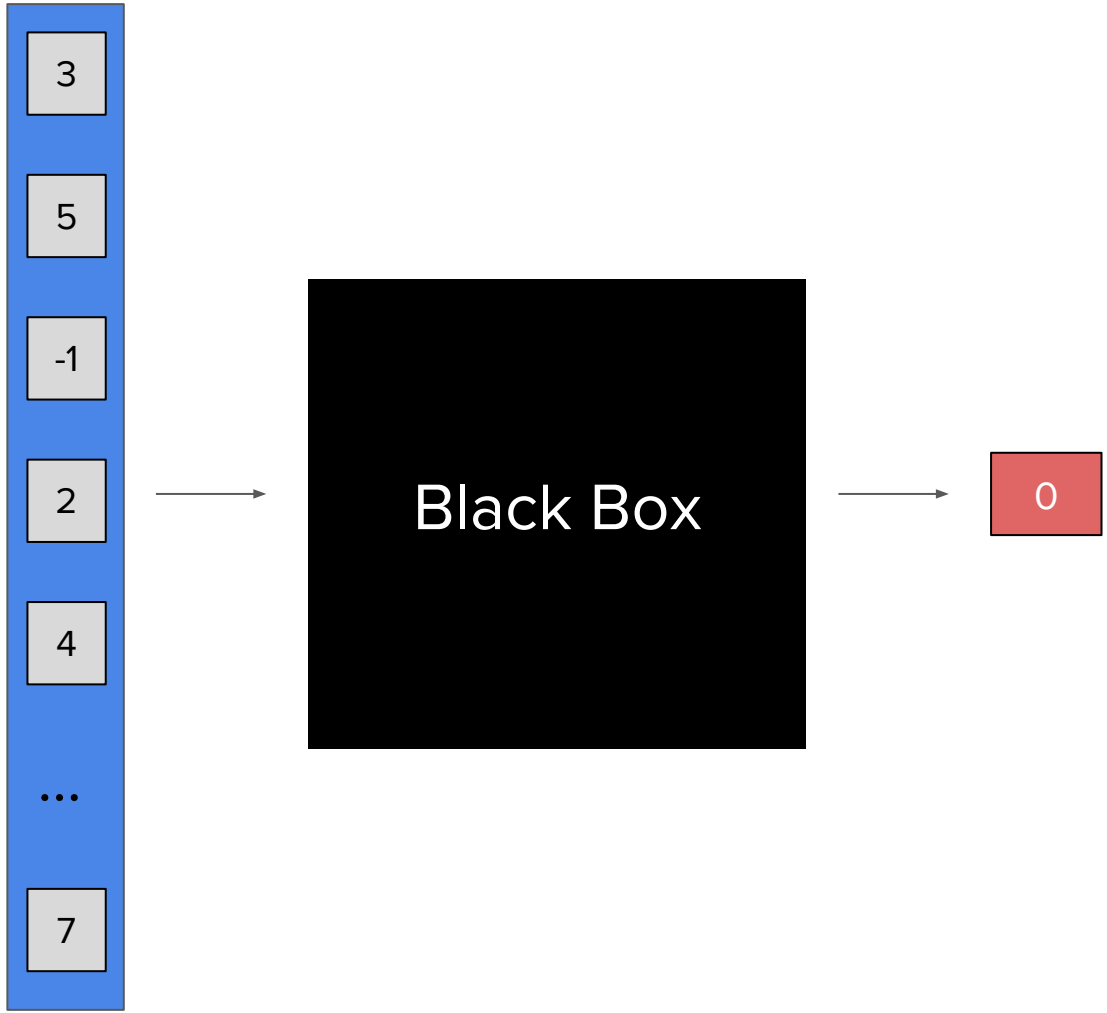
300-dimensional word vector

e.g. $v_{homemaker}$



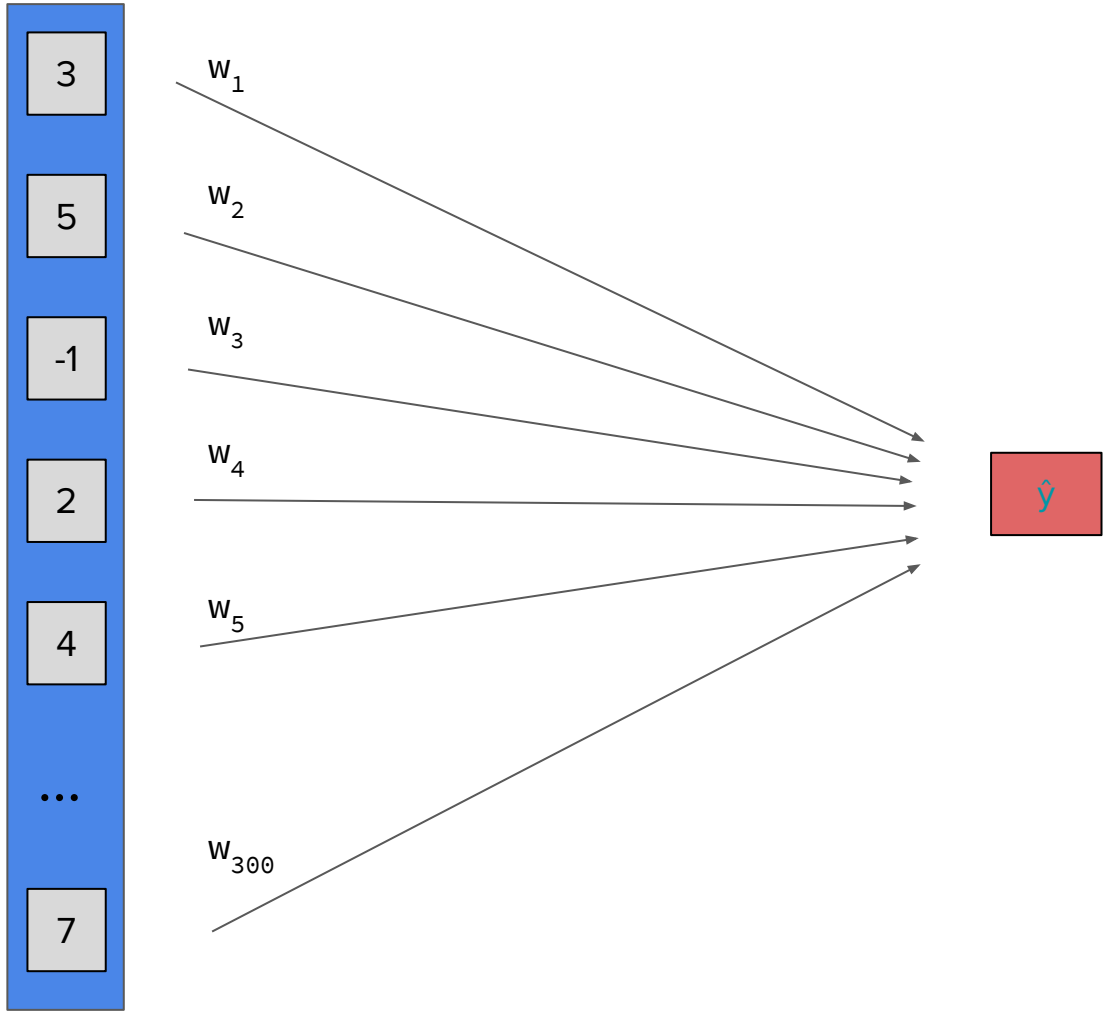
300-dimensional word vector

e.g. $v_{programmer}$



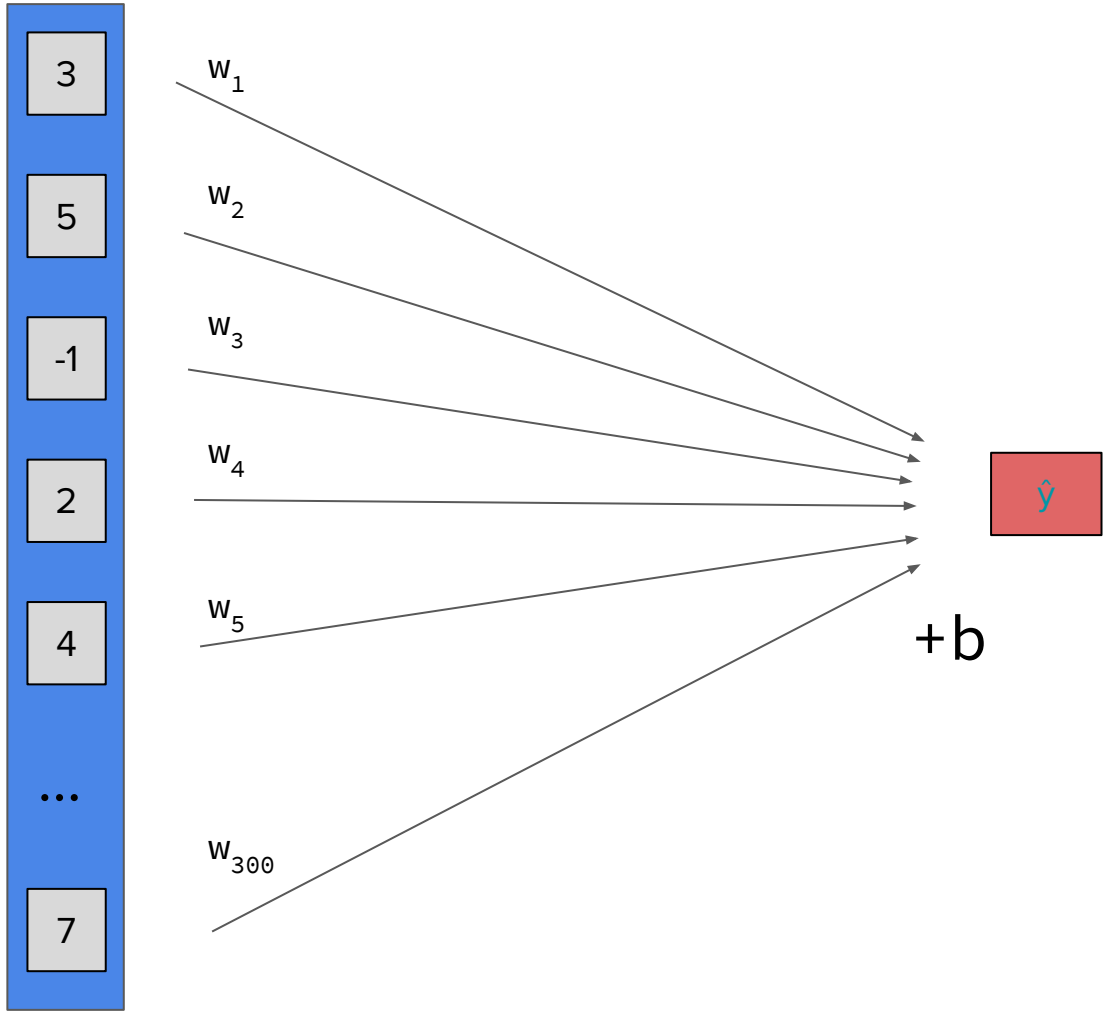
300-dimensional word vector

e.g. $v_{programmer}$

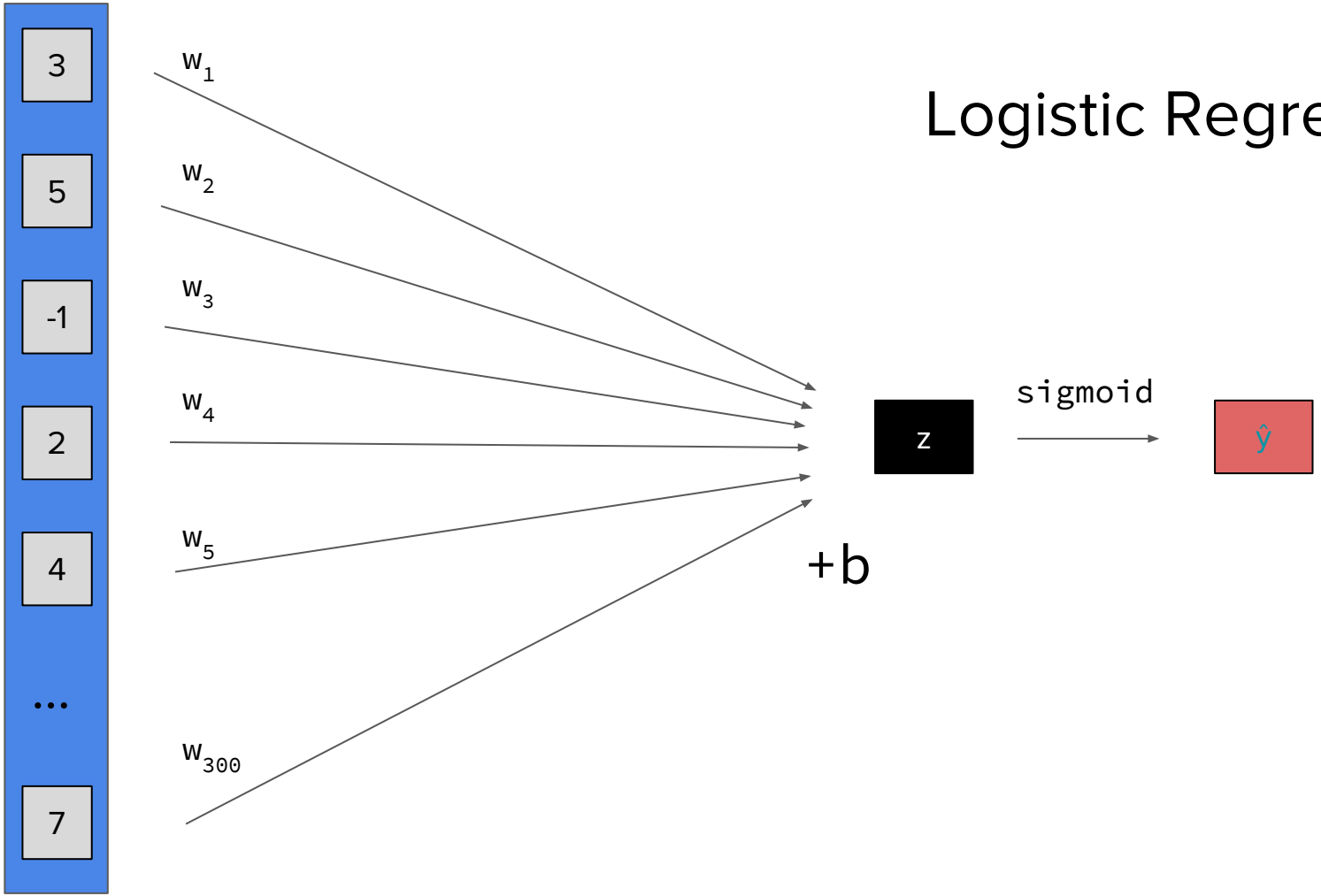


300-dimensional word vector

e.g. $v_{programmer}$

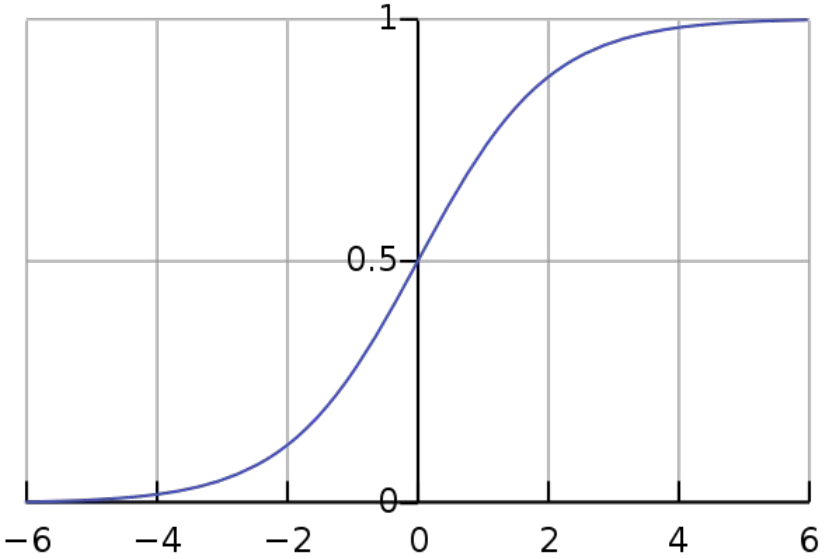


Logistic Regression

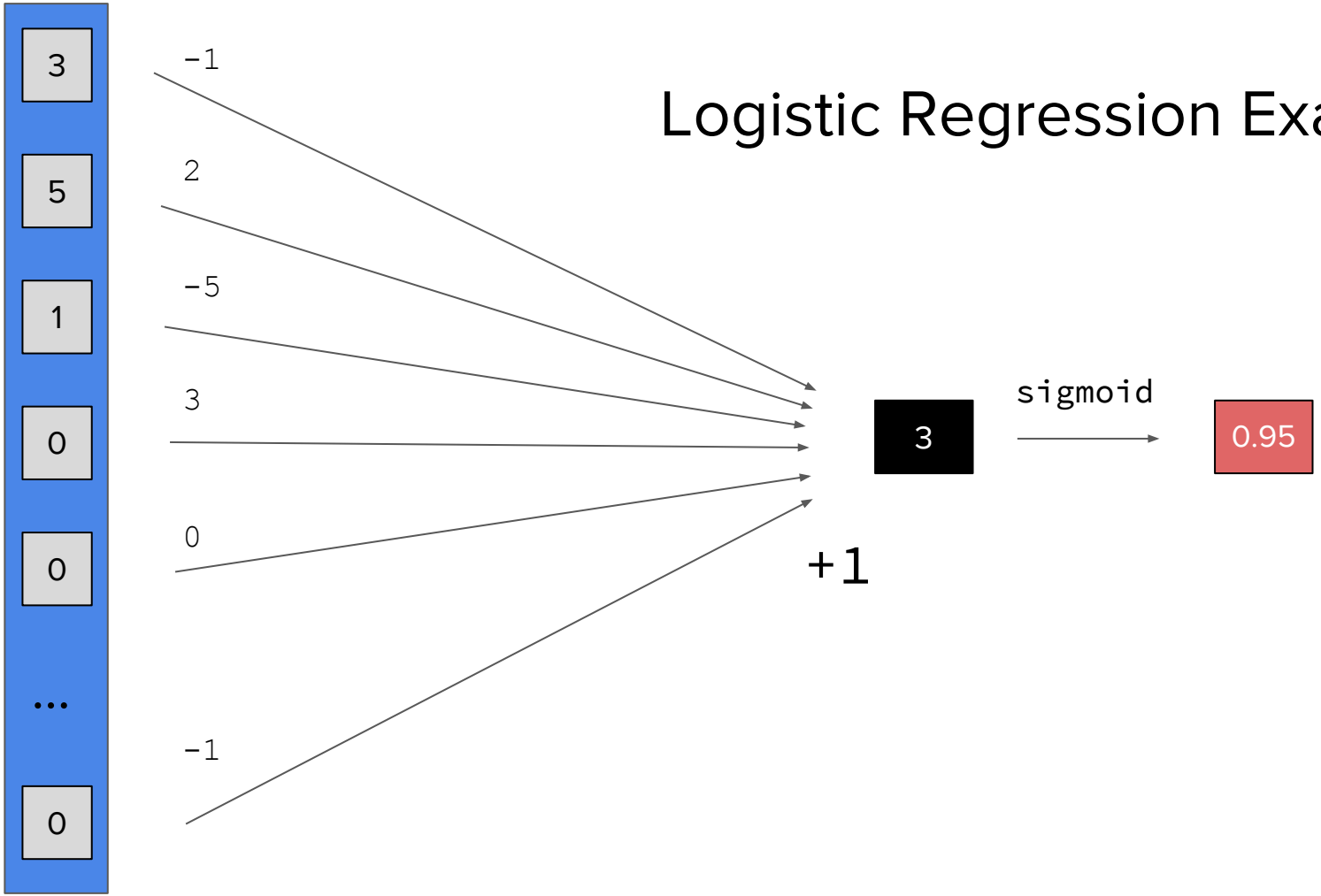


Sigmoid Function

$$\hat{y} = \frac{1}{1 + e^{-z}}$$



Logistic Regression Example



Jupyter Notebook Exercises: Part 2

Questions?

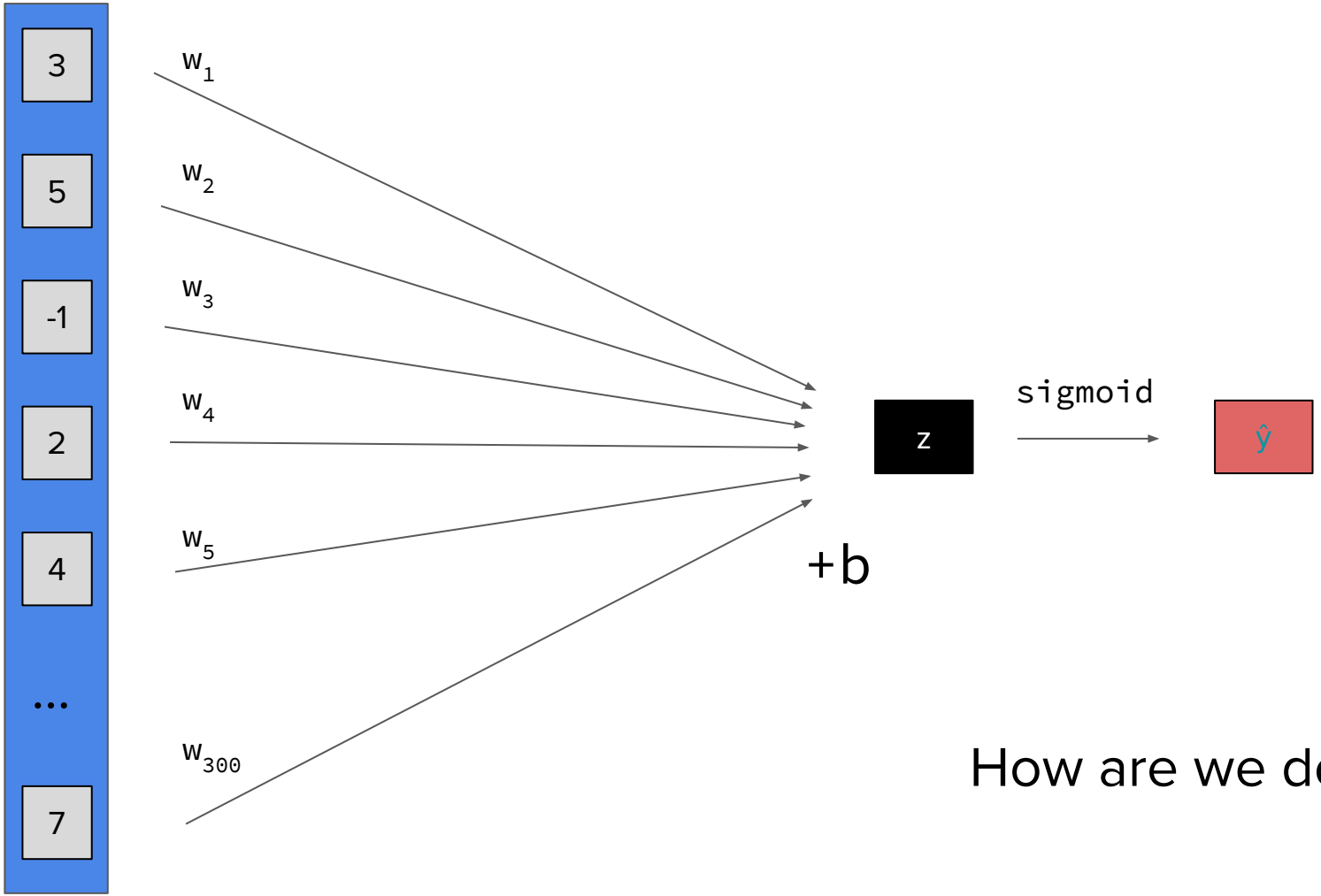
How do we choose *w* and *b*?

Why Train our Model?

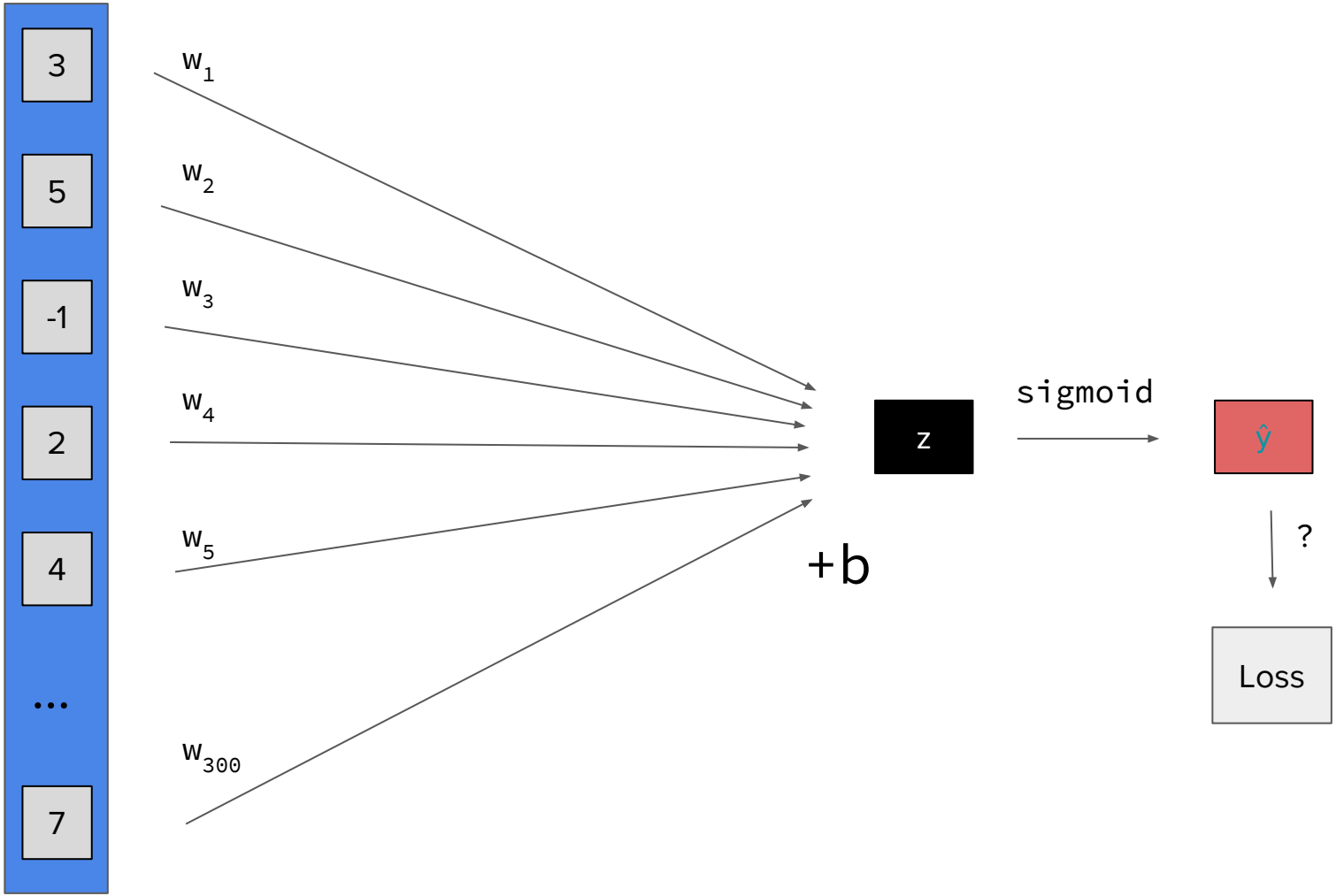
- For more complex models, you won't be able to guess the weights
- This is a more convincing demonstration of gender bias in word vectors

Training Data

Word (x)	**True Label (y)**
female	1
male	0
woman	1
man	0
...	...



How are we doing?


3
5
-1
2
4
...
7
w1
w2
w3
w4
w5
w300
+b
z
sigmoid
ŷ
?
Loss

What we've learned...

- AI for social good is important!
- How to build a linear model
- How to build a logistic regression model
- The basics of machine learning
- Word vectors contain alarming gender biases

Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings