Embeddings

Outline

- Motivation
- Categorical Embedding from Scratch
- Examples of Popular Embeddings
  - Rossmann Cat Embedding
  - Word2Vec
  - BERT
- Summary

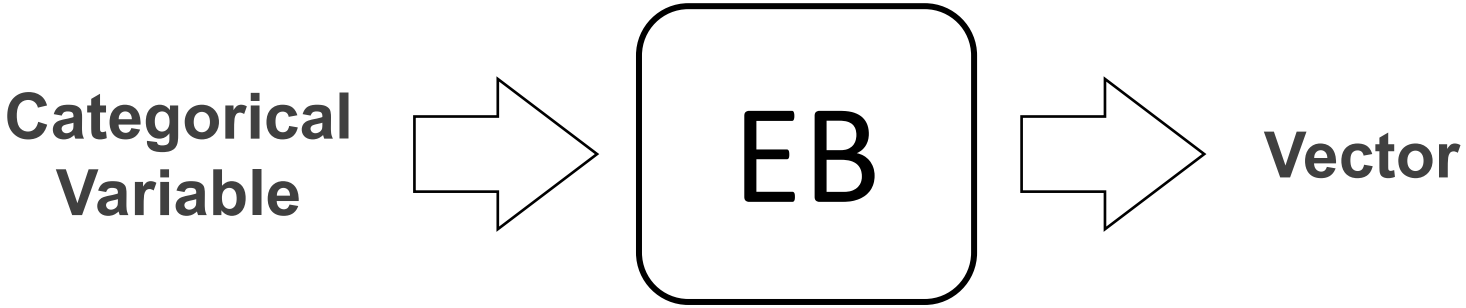
Motivation

Embeddings

- In short, embeddings maps categorical variables into vectors

Embeddings

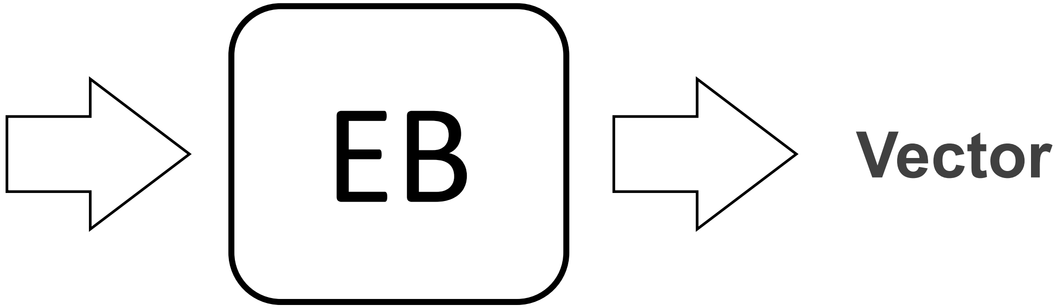
- In short, embeddings maps categorical variables into vectors



Embeddings

- In short, embeddings maps categorical variables into vectors

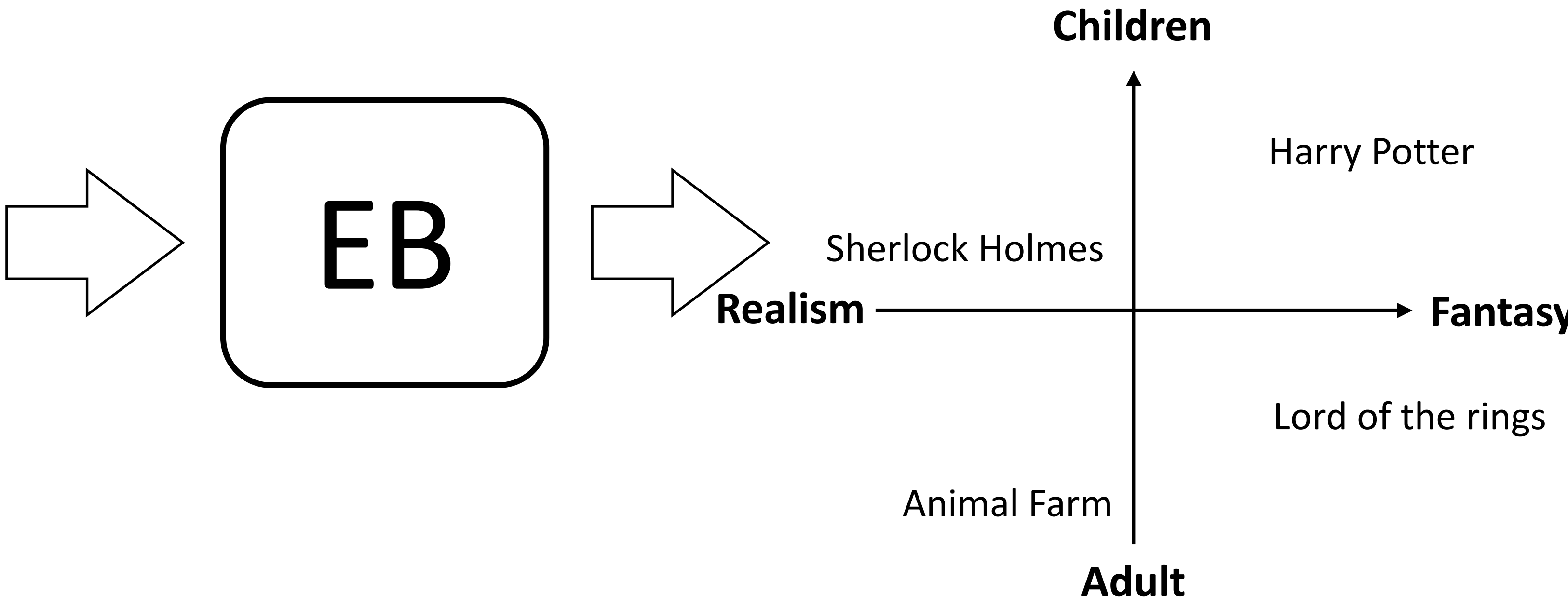
Book Title
Lord of the Rings
Harry Potter
Animal Farm
Sherlock Holmes



Embeddings

- In short, embeddings maps categorical variables into vectors

Book Title
Lord of the Rings
Harry Potter
Animal Farm
Sherlock Holmes

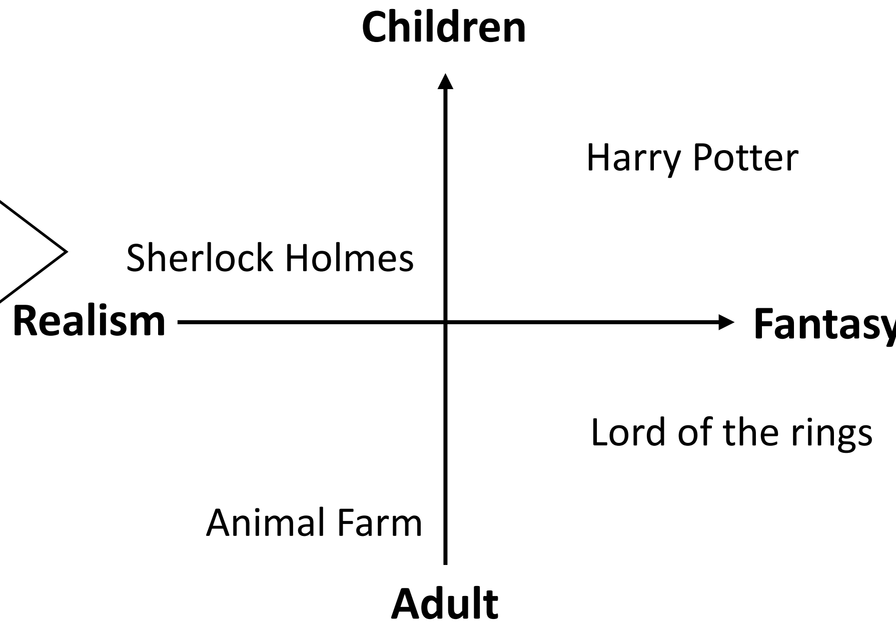


Embeddings

- In short, embeddings maps categorical variables into vectors

Book Title
Lord of the Rings
Harry Potter
Animal Farm
Sherlock Holmes

	Realism/ Fantasy	Children /Adult
Lord of the Rings	0.9	-0.3
Harry Potter	0.9	0.7
Animal Farm	-0.4	-0.8
Sherlock Holmes	-0.8	0.2

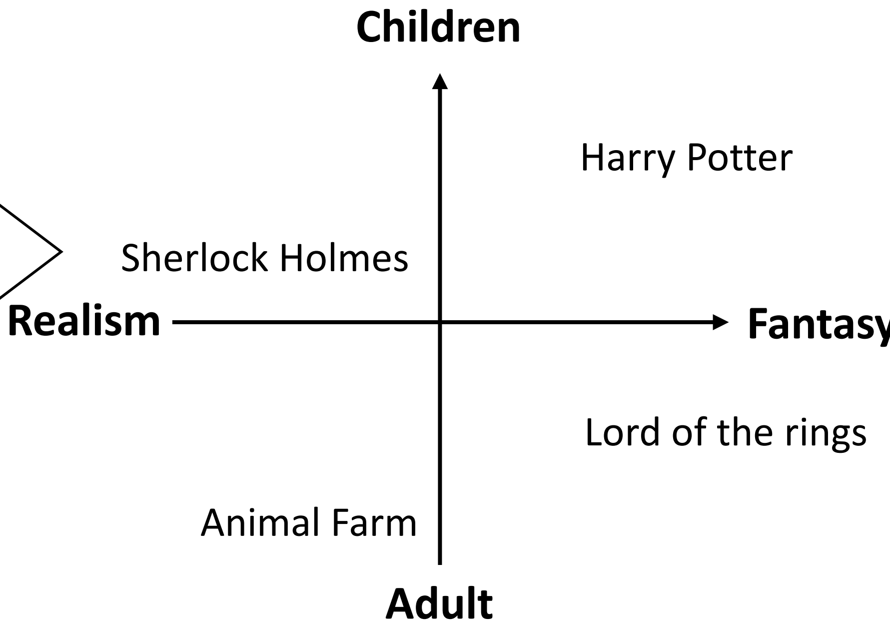


Embeddings

- In short, embeddings maps categorical variables into vectors

Book Title
Lord of the Rings
Harry Potter
Animal Farm
Sherlock Holmes

	Realism/ Fantasy	Children /Adult
Lord of the Rings	0.9	-0.3
Harry Potter	0.9	0.7
Animal Farm	-0.4	-0.8
Sherlock Holmes	-0.8	0.2



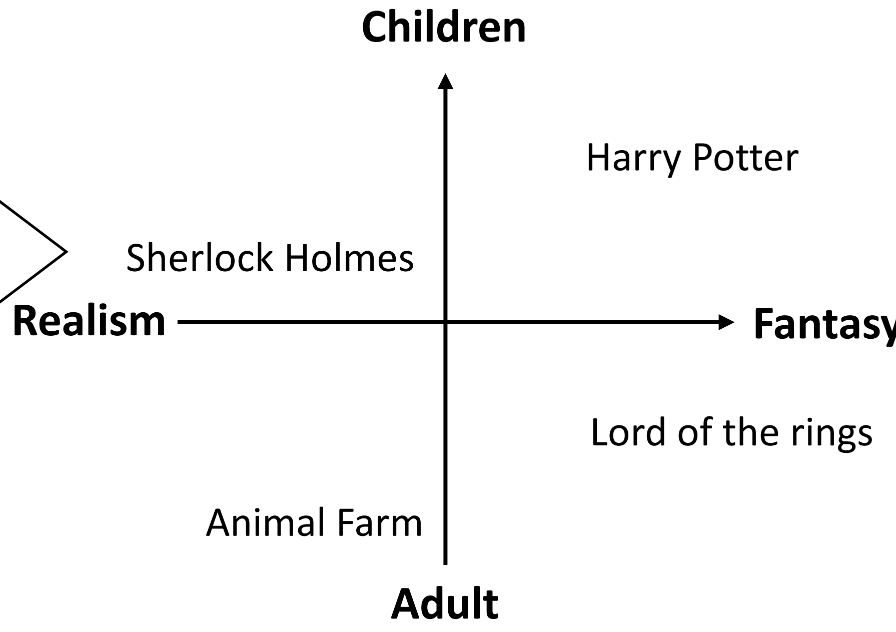
But how do we create such table?

Embeddings

- In short, embeddings maps categorical variables into vectors

Book Title
Lord of the Rings
Harry Potter
Animal Farm
Sherlock Holmes

	Realism/ Fantasy	Children /Adult
Lord of the Rings	0.9	-0.3
Harry Potter	0.9	0.7
Animal Farm	-0.4	-0.8
Sherlock Holmes	-0.8	0.2



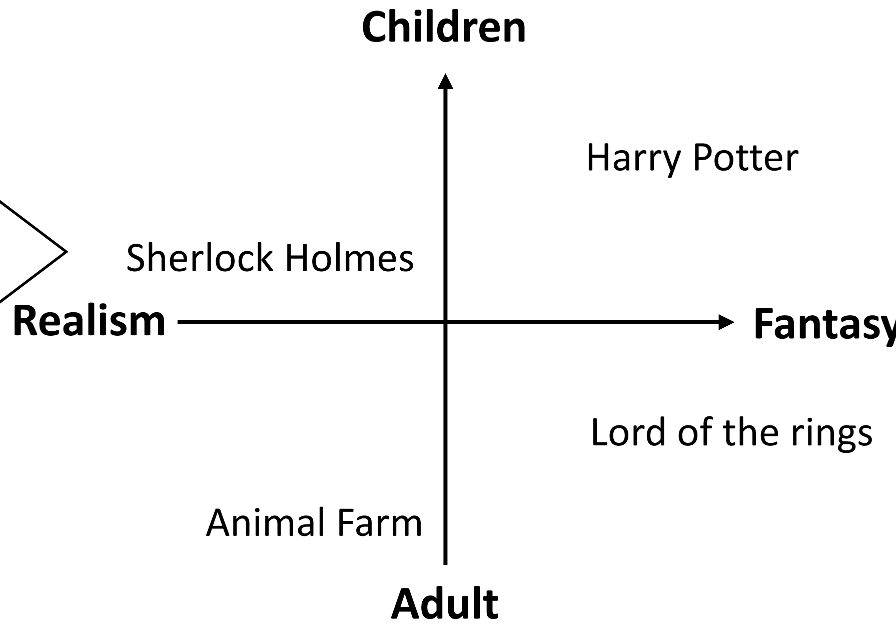
We need to set up a classification problem.

Embeddings

- In short, embeddings maps categorical variables into vectors

Book Title
Lord of the Rings
Harry Potter
Animal Farm
Sherlock Holmes

	Realism/ Fantasy	Children /Adult
Lord of the Rings	0.9	-0.3
Harry Potter	0.9	0.7
Animal Farm	-0.4	-0.8
Sherlock Holmes	-0.8	0.2



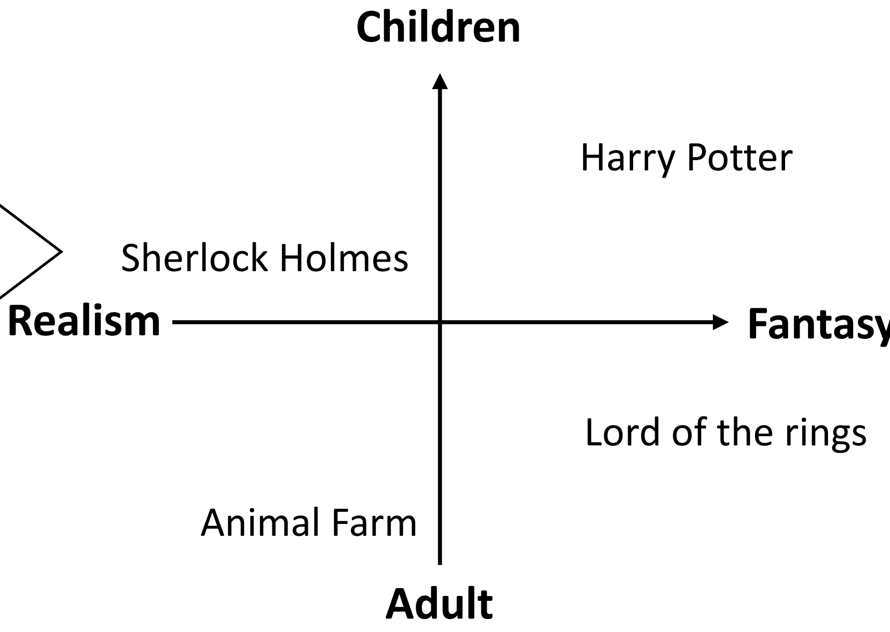
We need to set up a classification problem. Always!

Embeddings

- In short, embeddings maps categorical variables into vectors

Book Title
Lord of the Rings
Harry Potter
Animal Farm
Sherlock Holmes

	Realism/ Fantasy	Children /Adult
Lord of the Rings	0.9	-0.3
Harry Potter	0.9	0.7
Animal Farm	-0.4	-0.8
Sherlock Holmes	-0.8	0.2



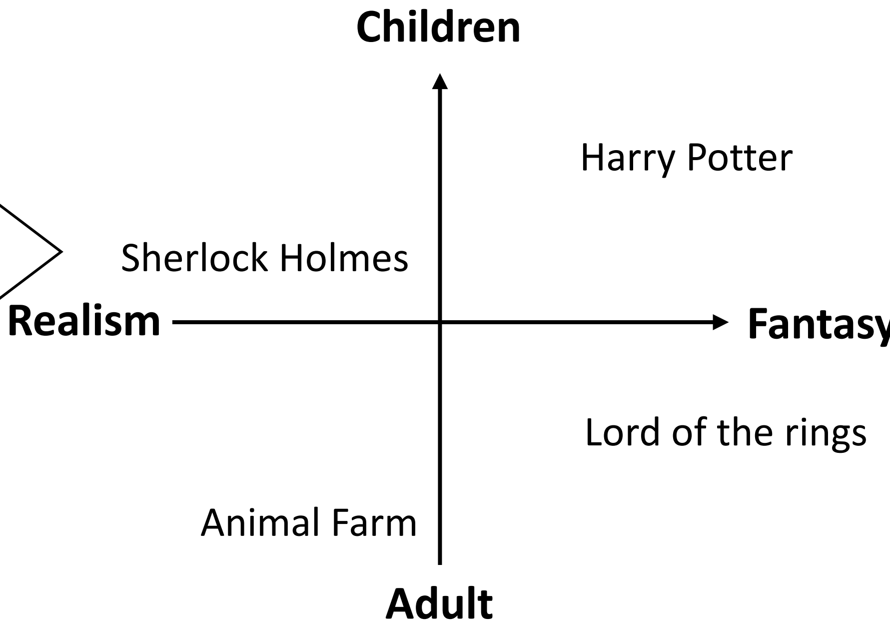
Embeddings are created using a supervised env

Embeddings

- In short, embeddings maps categorical variables into vectors

Book Title
Lord of the Rings
Harry Potter
Animal Farm
Sherlock Holmes

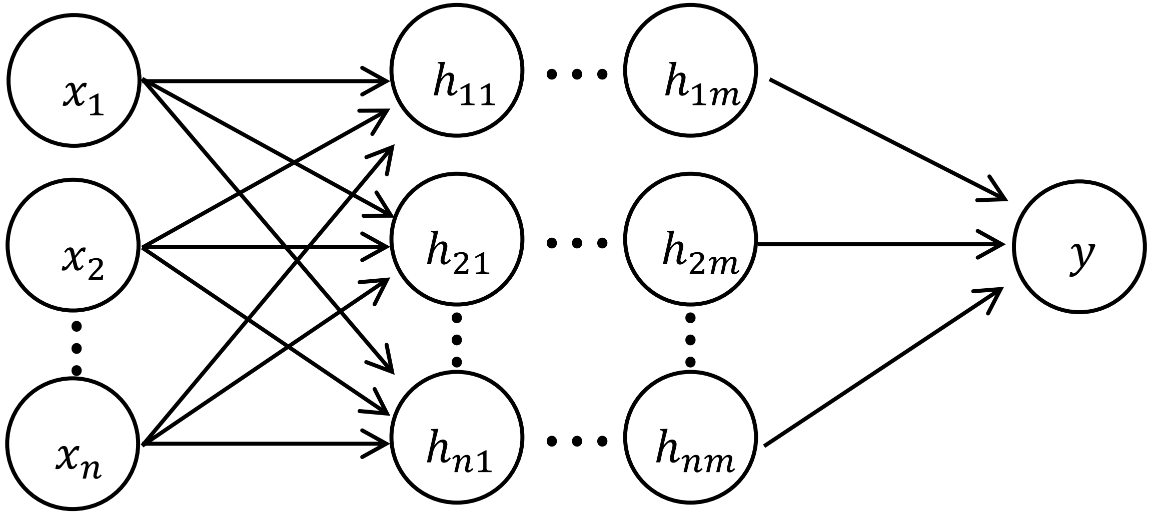
	Realism/ Fantasy	Children /Adult
Lord of the Rings	0.9	-0.3
Harry Potter	0.9	0.7
Animal Farm	-0.4	-0.8
Sherlock Holmes	-0.8	0.2



But can be used for unsupervised problems!

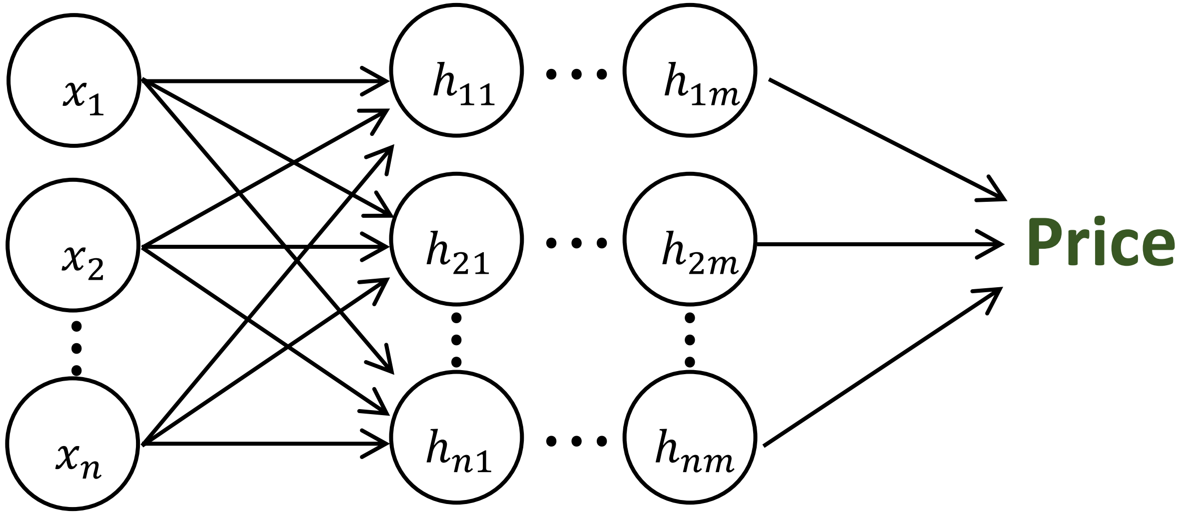
Embedding for Categorical Encoding from Scratch

Embeddings



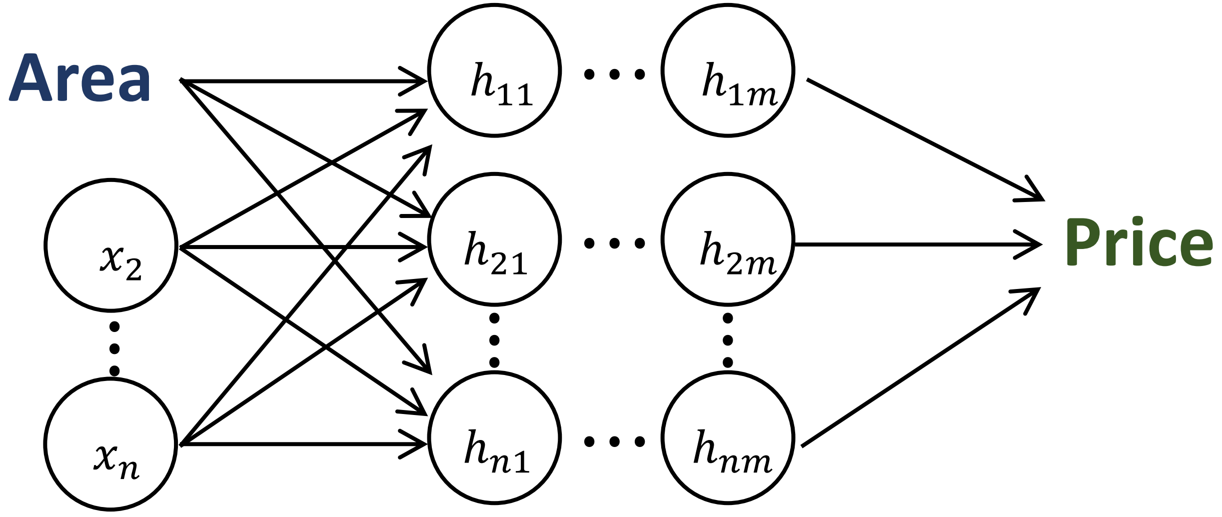
Embeddings

			Price
			500k
			600k
			900k
			1200k
			…



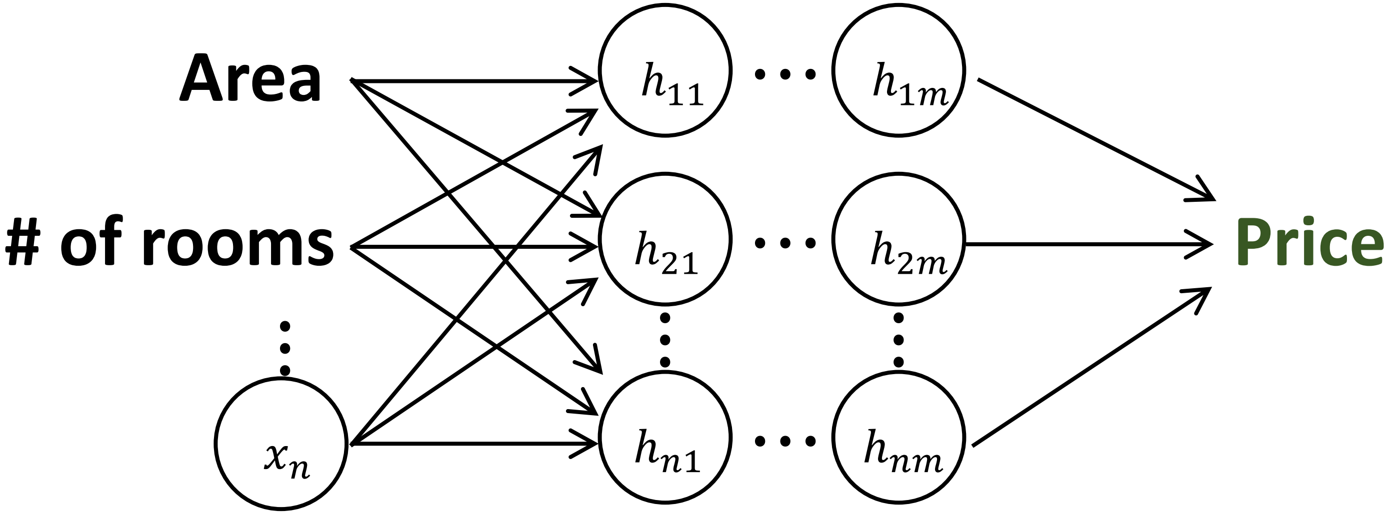
Embeddings

		Area(m2)	Price
		61	500k
		72	600k
		83	900k
		91	1200k
		…	…



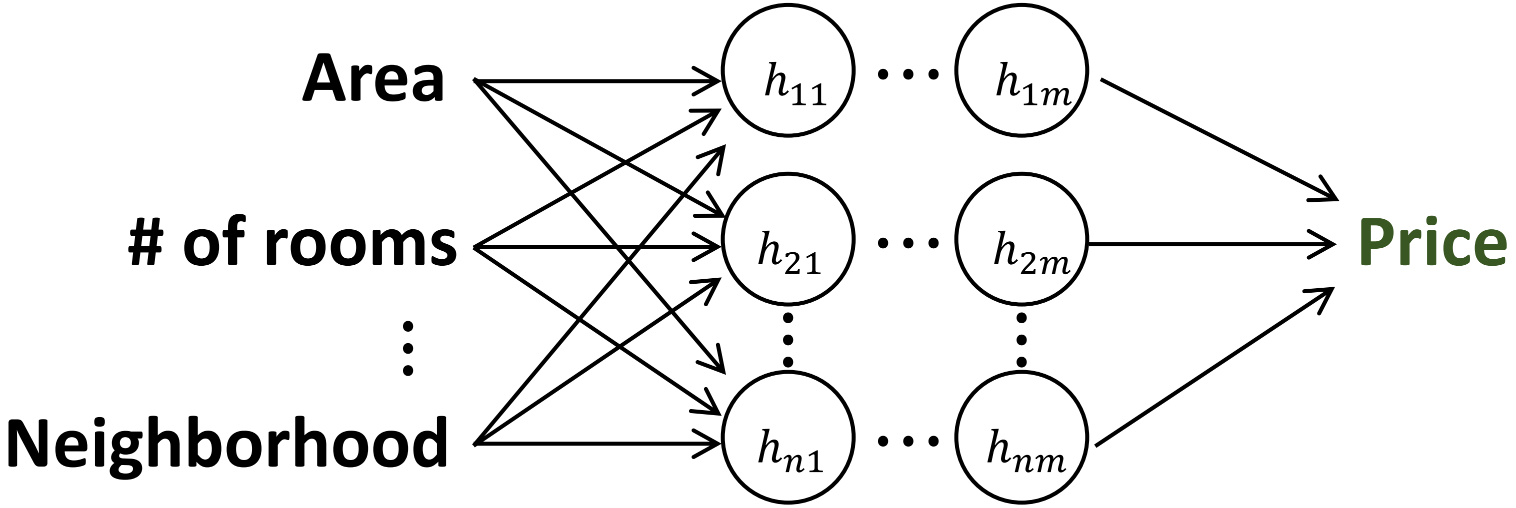
Embeddings

	rooms	Area(m2)	Price
	4	61	500k
	5	72	600k
	6	83	900k
	7	91	1200k
	…	…	…



Embeddings

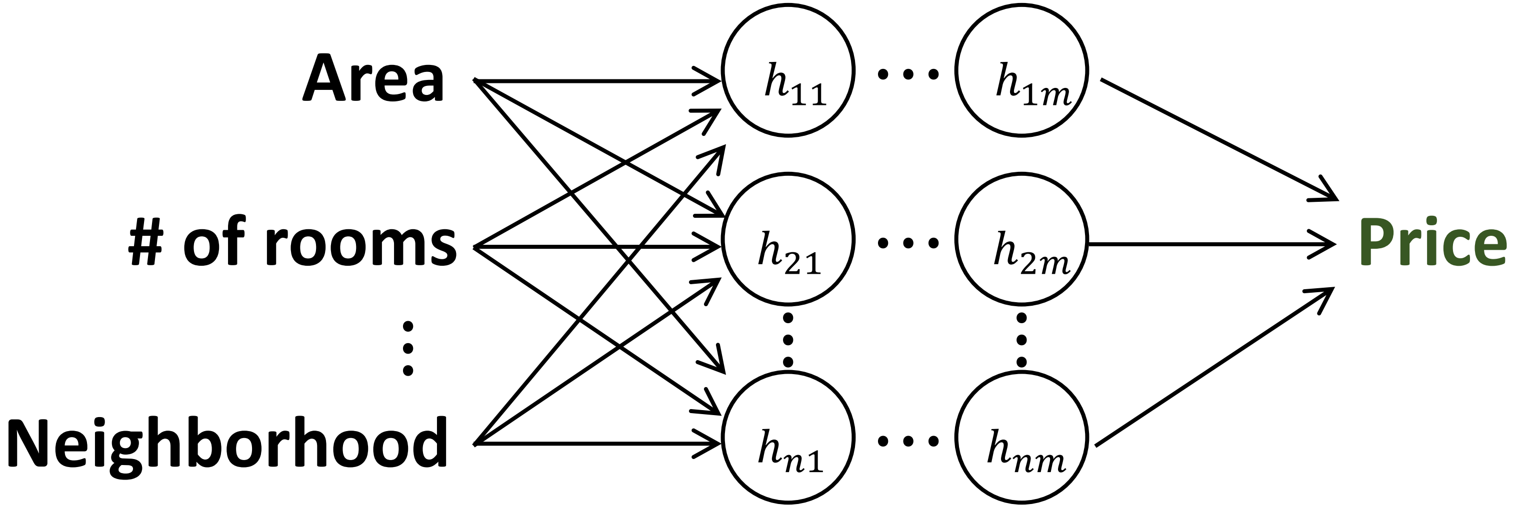
Nbhd	rooms	Area(m2)	Price
Tender	4	61	500k
Civic	5	72	600k
FinDistr	6	83	900k
Soma	7	91	1200k
…	…	…	…



Embeddings

!

Nbhd	rooms	Area(m2)	Price
Tender	4	61	500k
Civic	5	72	600k
FinDistr	6	83	900k
Soma	7	91	1200k
…	…	…	…

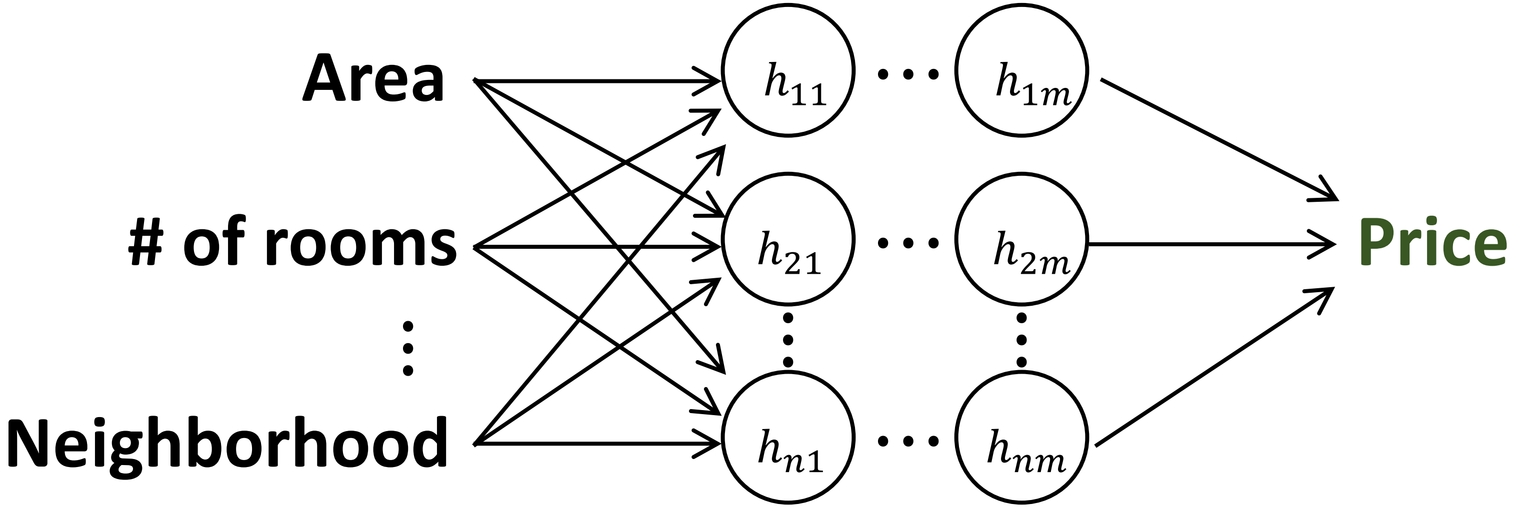


Embeddings

Nbhd	rooms	Area(m2)	Price
Tender	4	61	500k
Civic	5	72	600k
FinDistr	6	83	900k
Soma	7	91	1200k
…	…	…	…

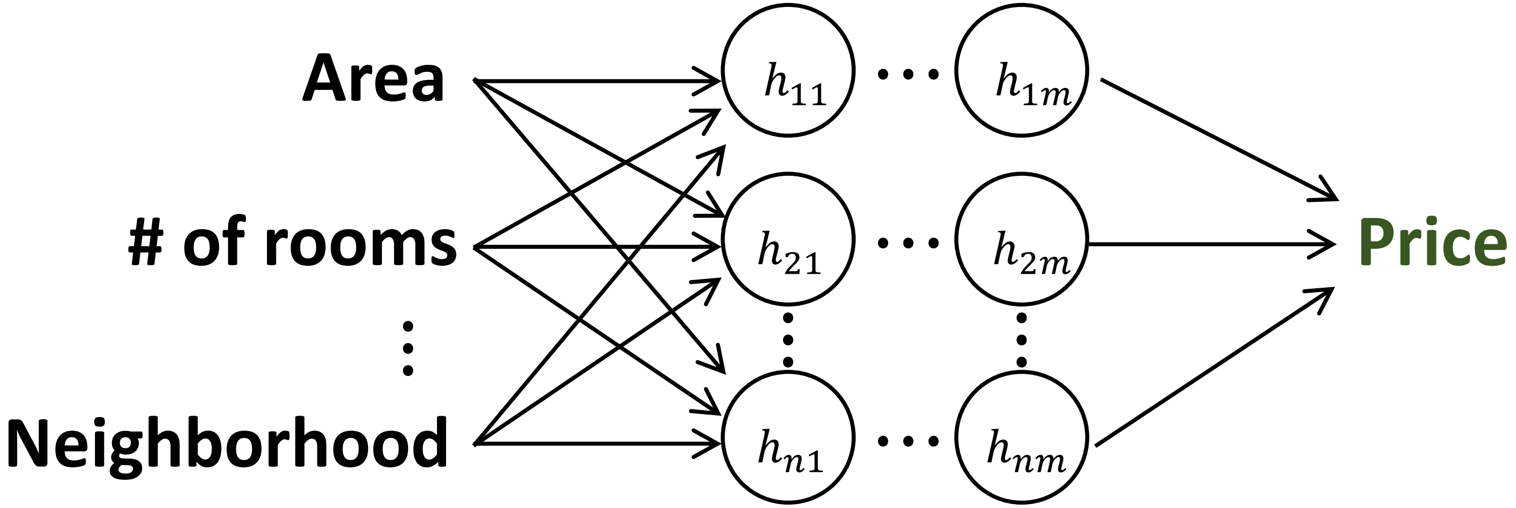
!

Inputs need to be numeric



Embeddings

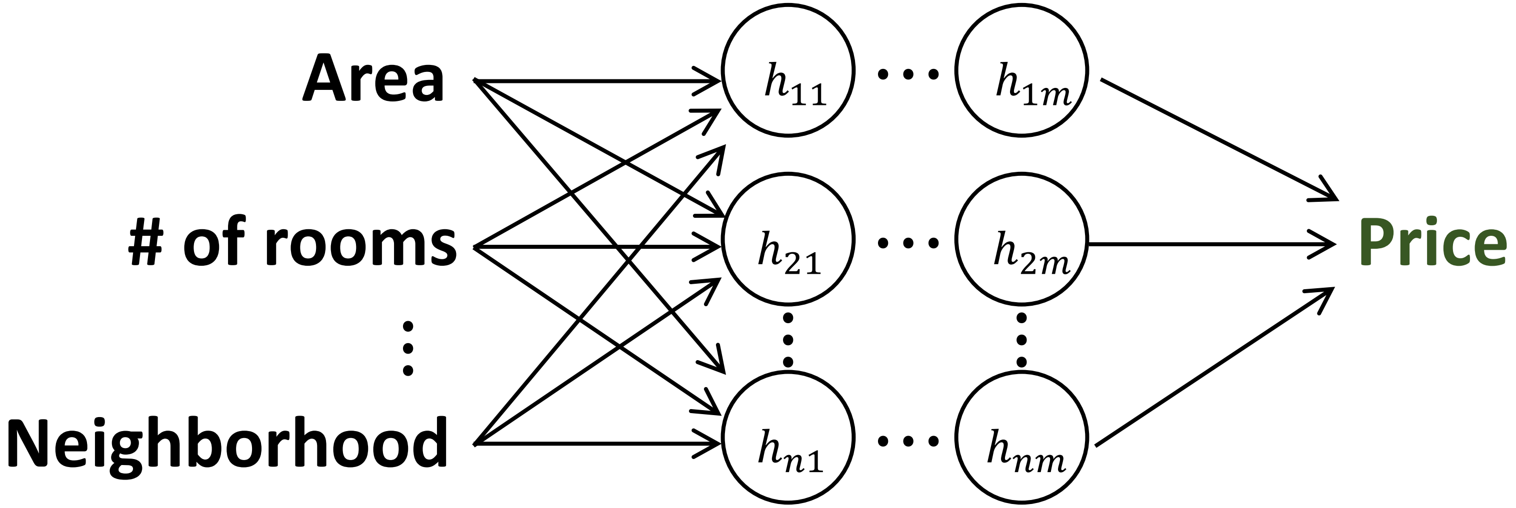
Nbhd	rooms	Area(m2)	Price
0	4	61	500k
1	5	72	600k
2	6	83	900k
3	7	91	1200k
…	…	…	…



Embeddings

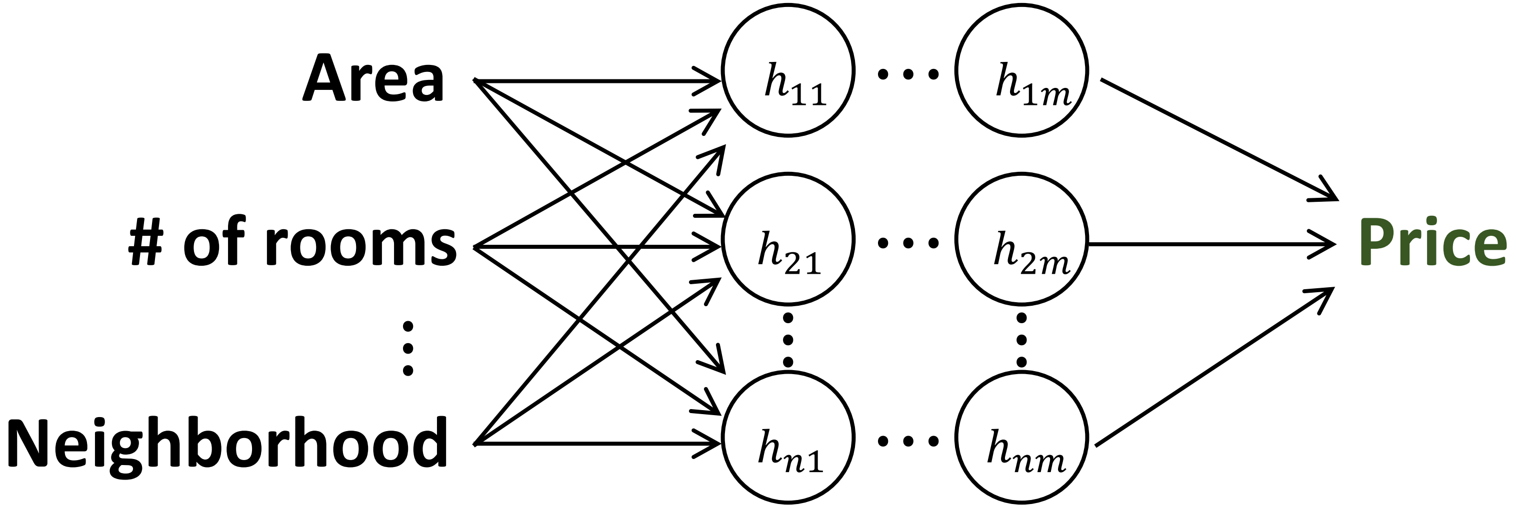
- However, there’s no proximity relationship

Nbhd	rooms	Area(m2)	Price
0	4	61	500k
1	5	72	600k
2	6	83	900k
3	7	91	1200k
…	…	…	…



Embeddings

Nbhd	rooms	Area(m2)	Price
Tender	4	61	500k
Civic	5	72	600k
FinDistr	6	83	900k
Soma	7	91	1200k
…	…	…	…



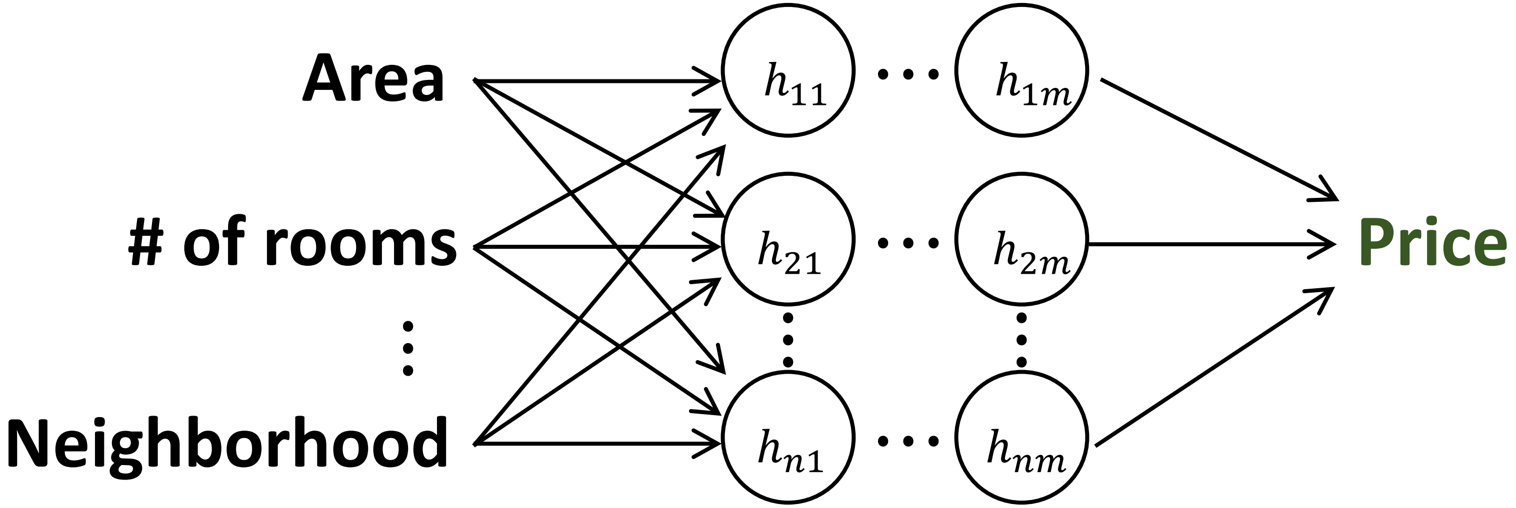
Embeddings

- Another solution – One-Hot Encoding

Nbhd	rooms	Area(m2)	Price
Tender	4	61	500k
Civic	5	72	600k
FinDistr	6	83	900k
Soma	7	91	1200k
…	…	…	…

One-Hot Encoding

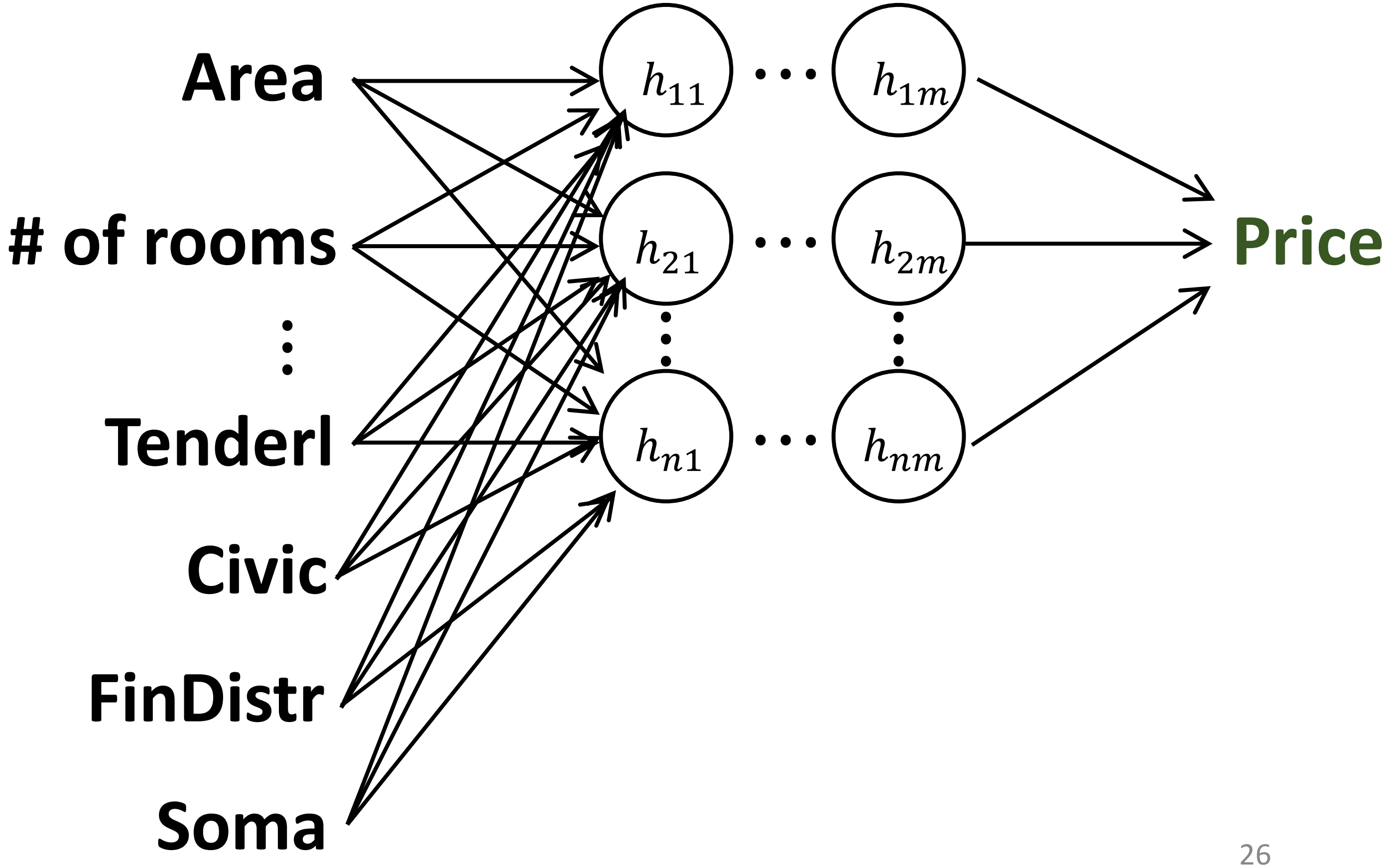
Tenderl	Civic	FinDistr	Soma
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1



Embeddings

- Another solution – One-Hot Encoding

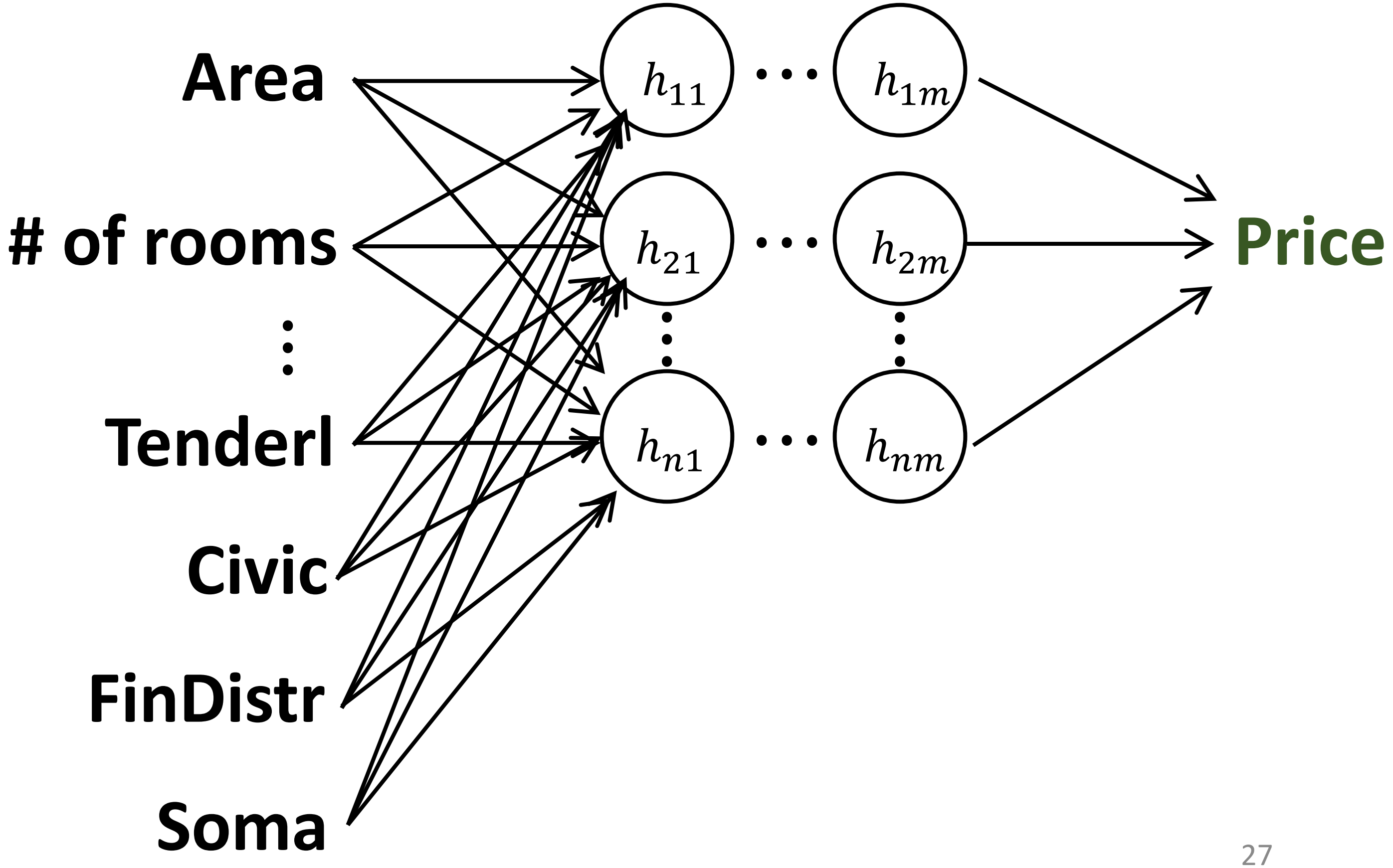
Tenderl	Civic	FinDistr	Soma	rooms	Area)	Price
1	0	0	0	4	61	500k
0	1	0	0	5	72	600k
0	0	1	0	6	83	900k
0	0	0	1	7	91	1200k
…	…	…	…	…	…	…



Embeddings

- However, it creates very sparse matrices
  - Unable to correlate similar categories

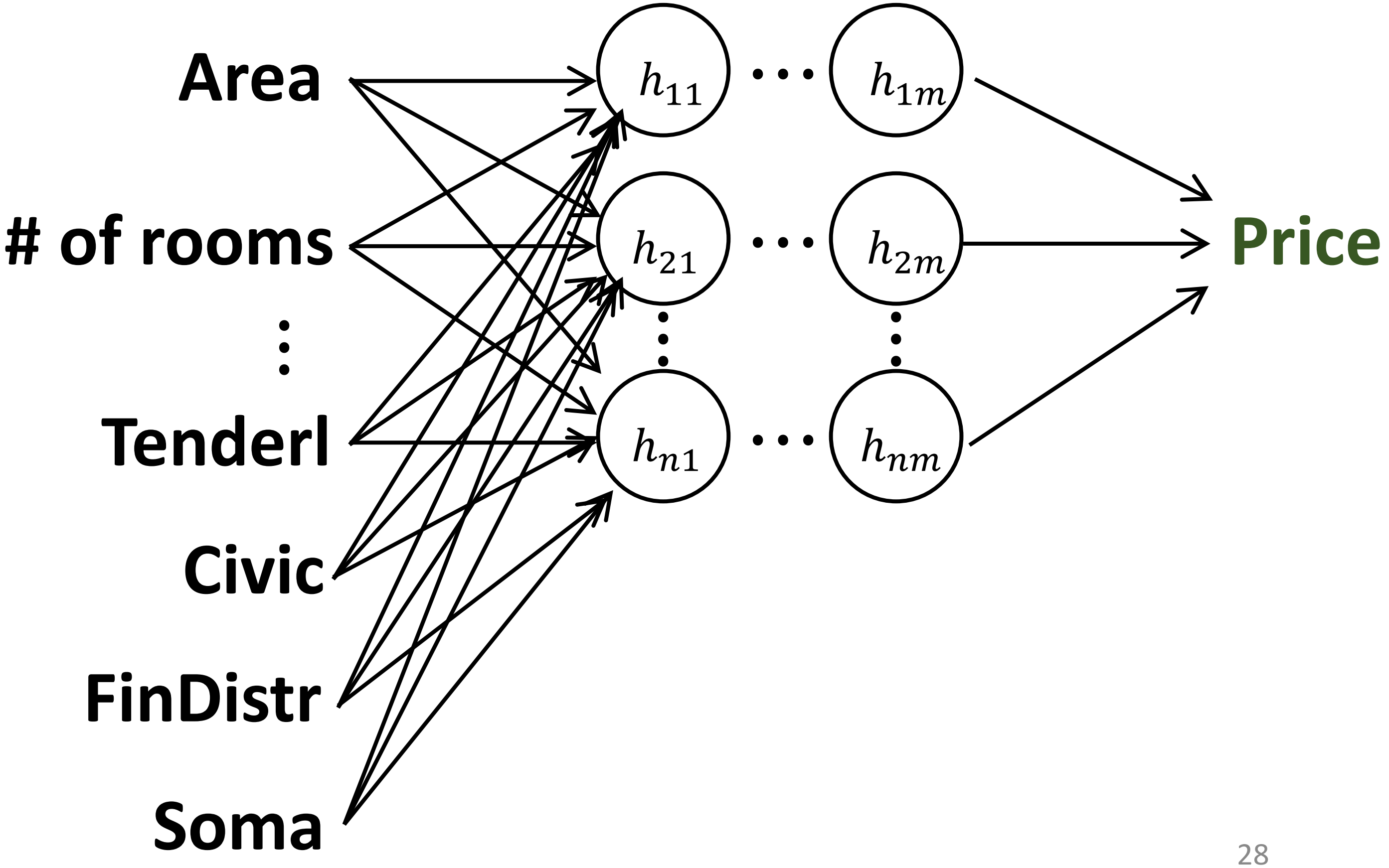
Tenderl	Civic	FinDistr	Soma	rooms	Area)	Price
1	0	0	0	4	61	500k
0	1	0	0	5	72	600k
0	0	1	0	6	83	900k
0	0	0	1	7	91	1200k
…	…	…	…	…	…	…



Embeddings

- However, it creates very sparse matrices
  - Unable to correlate similar categories

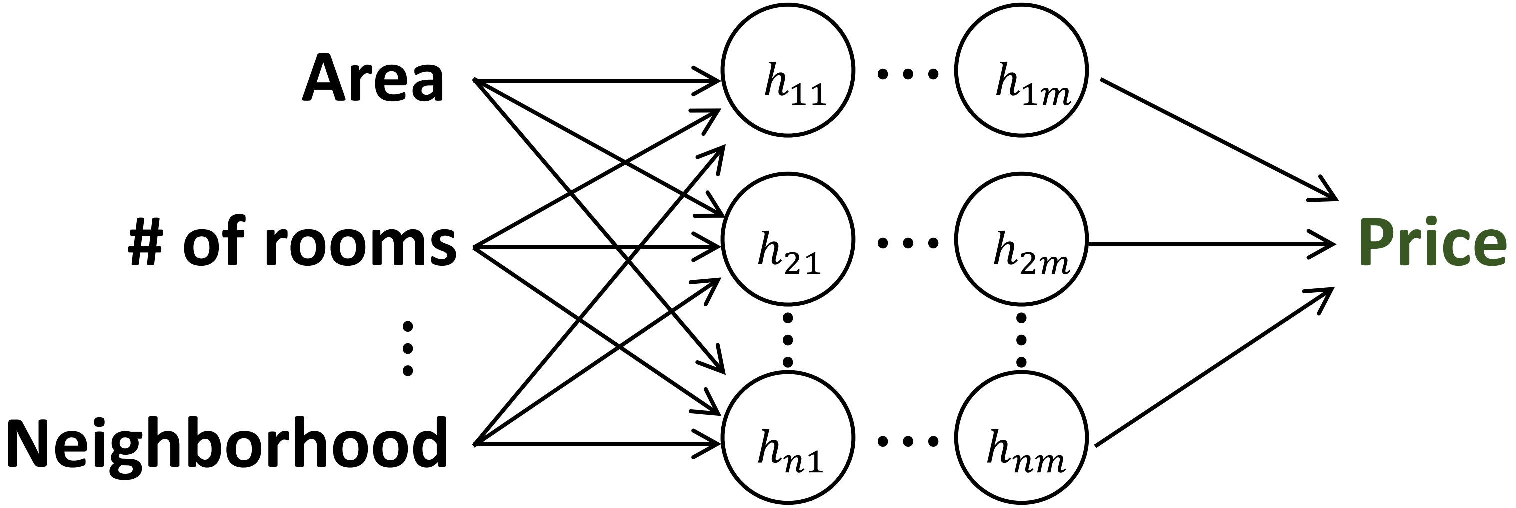
Tenderl	Civic	FinDistr	Soma	rooms	Area)	Price
1	0	0	0	4	61	500k
0	1	0	0	5	72	600k
0	0	1	0	6	83	900k
0	0	0	1	7	91	1200k
…	…	…	…	…	…	…



HAVE YOU EVER BEEN TO SF?

Embeddings

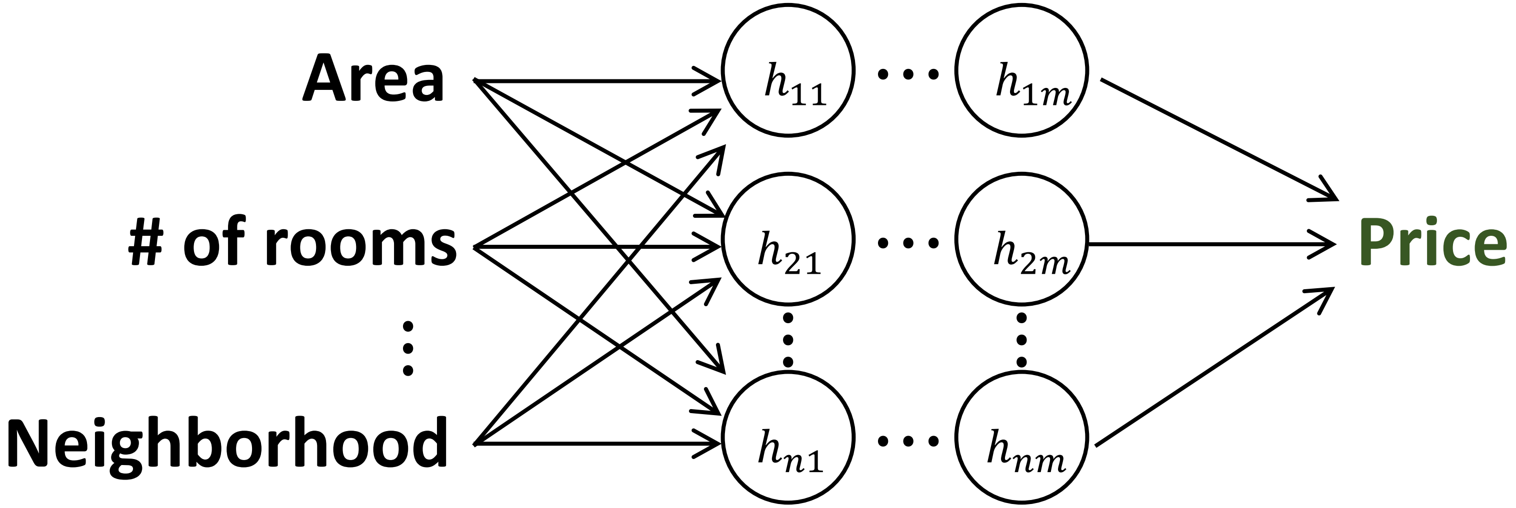
Nbhd	rooms	Area(m2)	Price
Tender	4	61	500k
Civic	5	72	600k
FinDistr	6	83	900k
Soma	7	91	1200k
…	…	…	…



Embeddings

- Solution: Embeddings

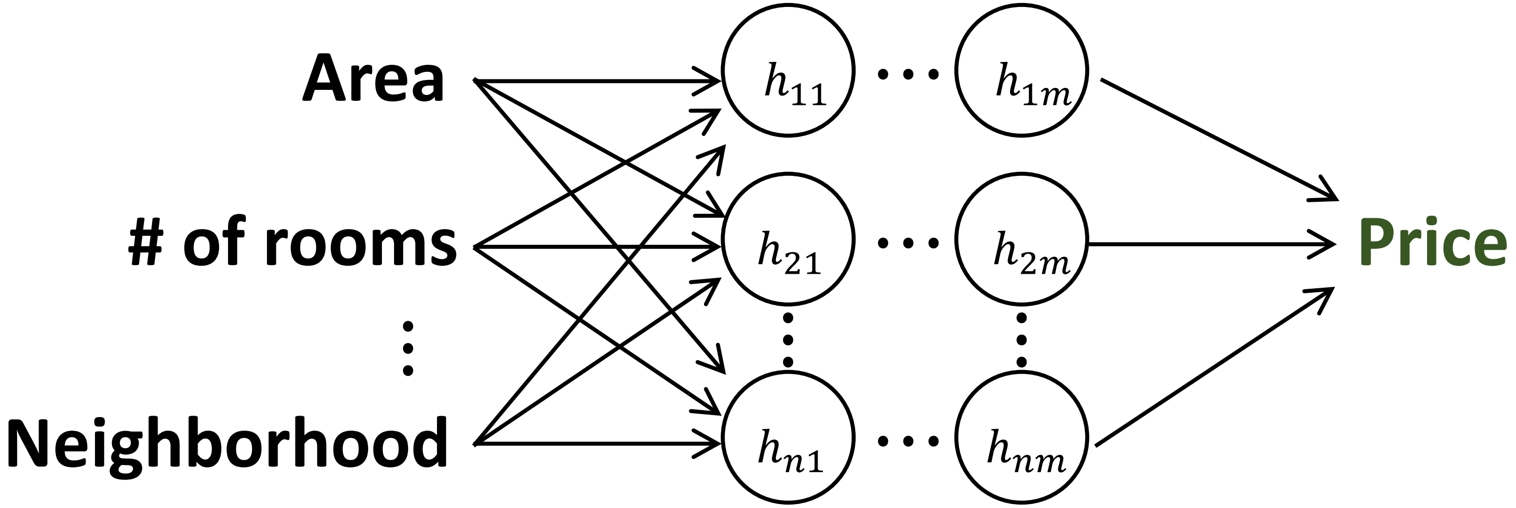
Nbhd	rooms	Area(m2)	Price
Tender	4	61	500k
Civic	5	72	600k
FinDistr	6	83	900k
Soma	7	91	1200k
…	…	…	…



Embeddings

- Solution: Embeddings
  - First step: replace values by their indexes

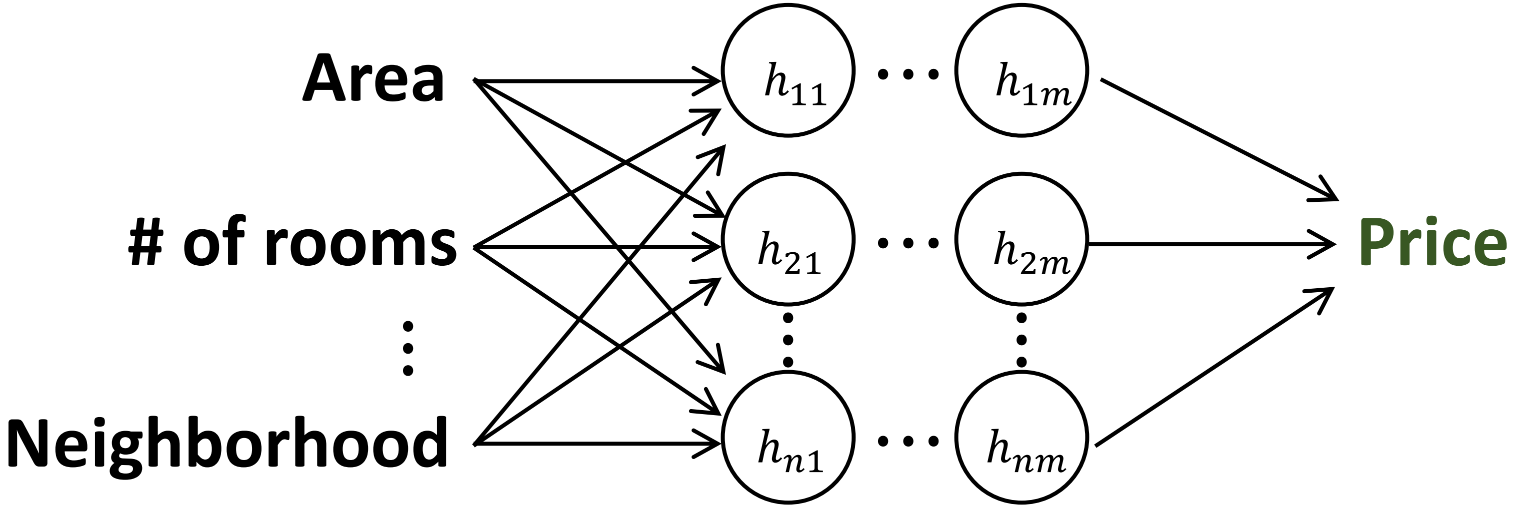
Nbhd	rooms	Area(m2)	Price
Tender	4	61	500k
Civic	5	72	600k
FinDistr	6	83	900k
Soma	7	91	1200k
…	…	…	…



Embeddings

- Solution: Embeddings
  - First step: replace values by their indexes

Nbhd	rooms	Area(m2)	Price
0	4	61	500k
1	5	72	600k
2	6	83	900k
3	7	91	1200k
…	…	…	…

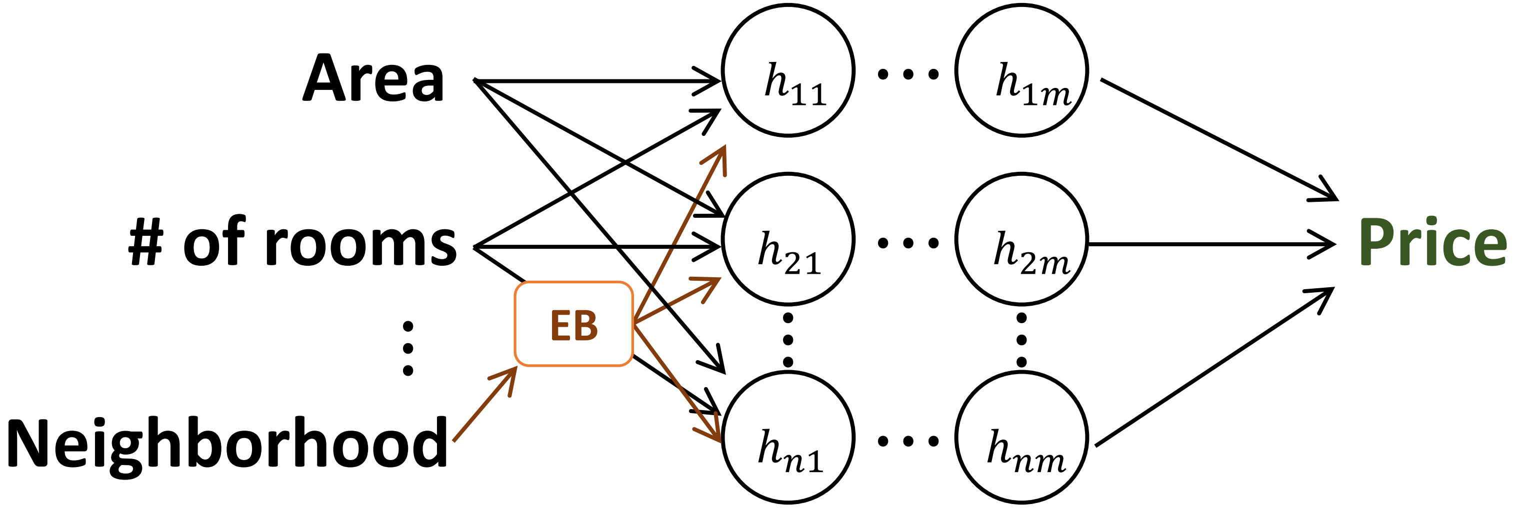



```
neibhd_to_idx={0:'Tender',1:'Civic',2:'FDist',3:'Soma'}
```

Embeddings

- Solution: Embeddings
  - Use an embedding layer instead of feeding directly

Nbhd	rooms	Area(m2)	Price
0	4	61	500k
1	5	72	600k
2	6	83	900k
3	7	91	1200k
…	…	…	…

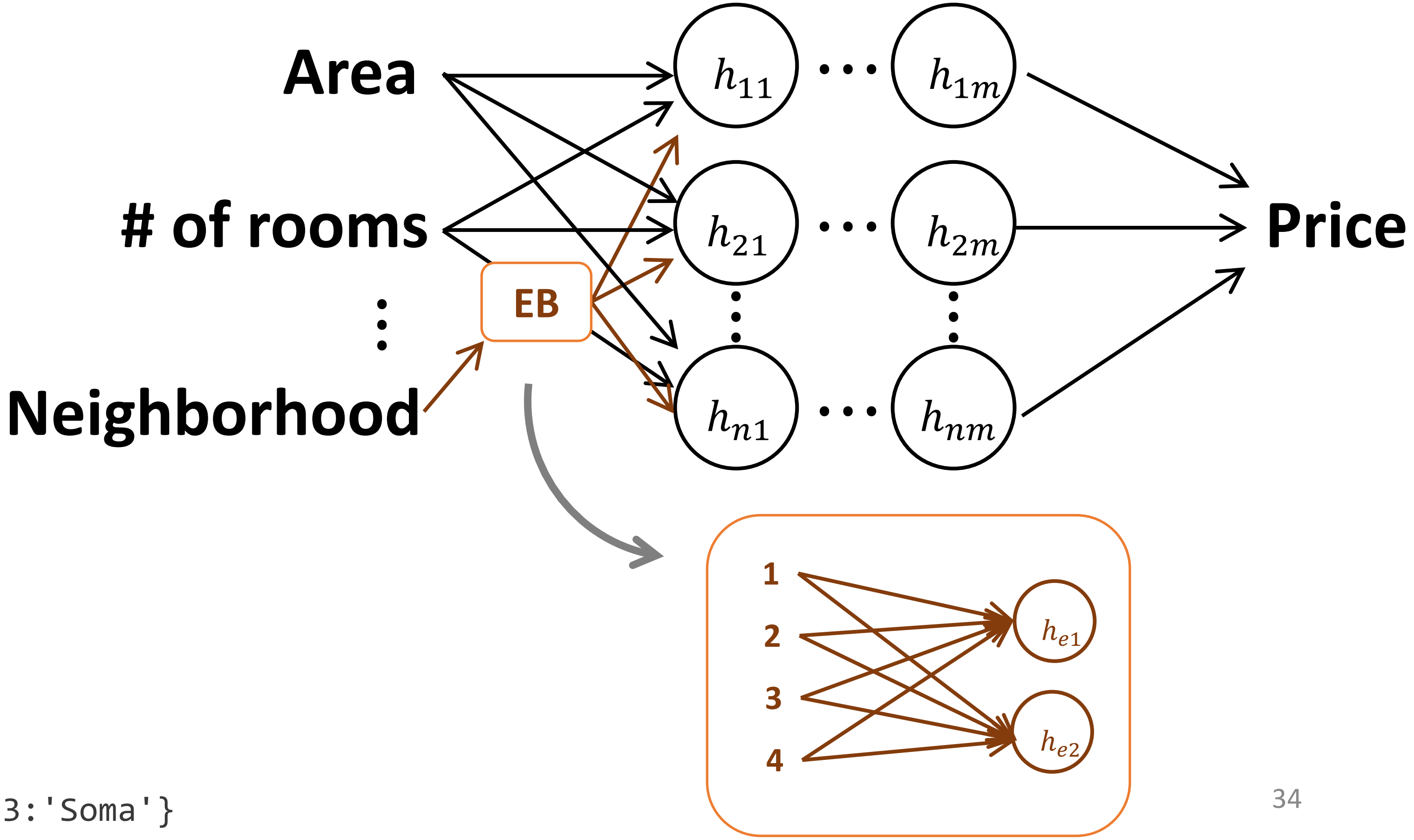



```
neibhd_to_idx={0:'Tender',1:'Civic',2:'FDist',3:'Soma'}
```

Embeddings

- Solution: Embeddings
  - Let’s take a closer look…

Nbhd	rooms	Area(m2)	Price
0	4	61	500k
1	5	72	600k
2	6	83	900k
3	7	91	1200k
…	…	…	…

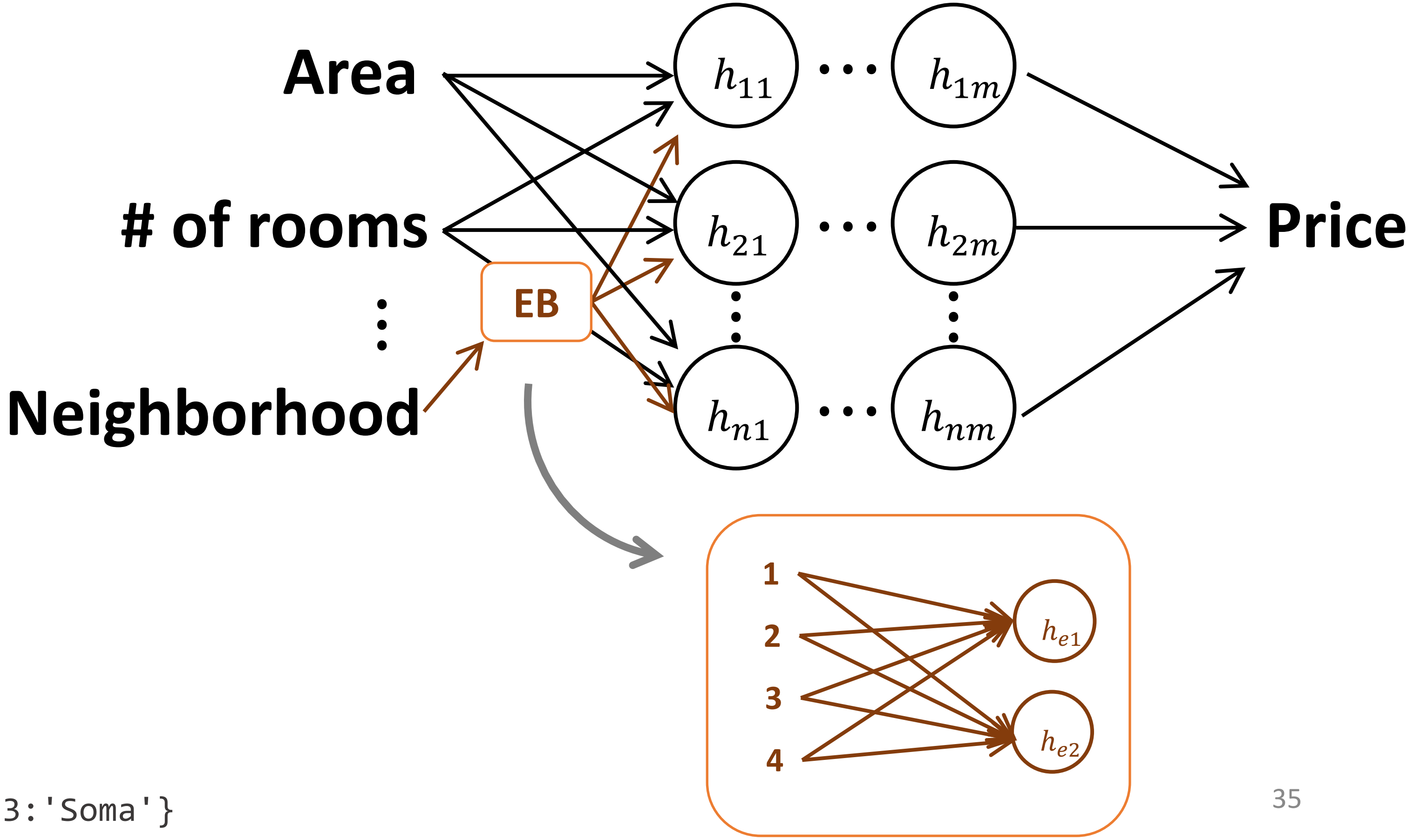



```
neibhd_to_idx={0:'Tender',1:'Civic',2:'FDist',3:'Soma'}
```

Embeddings

- Solution: Embeddings
  - At first, they look like a densely connected network
  - The embedding dimension (# of hidden layers) is chosen manually

Nbhd	rooms	Area(m2)	Price
0	4	61	500k
1	5	72	600k
2	6	83	900k
3	7	91	1200k
…	…	…	…

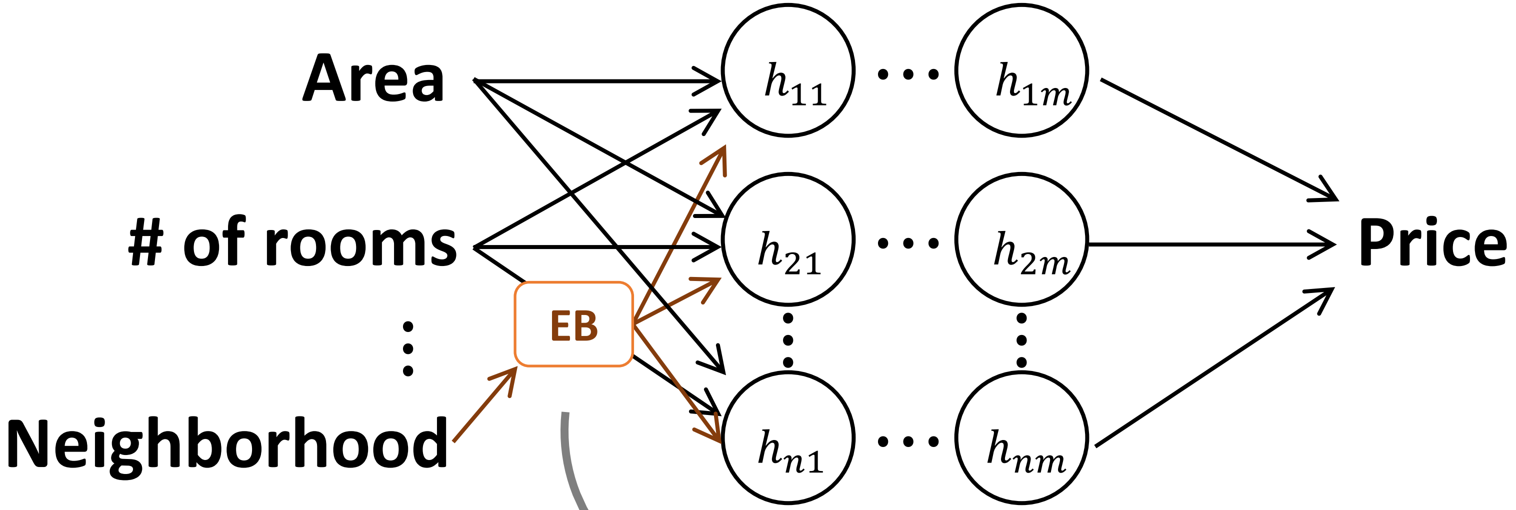


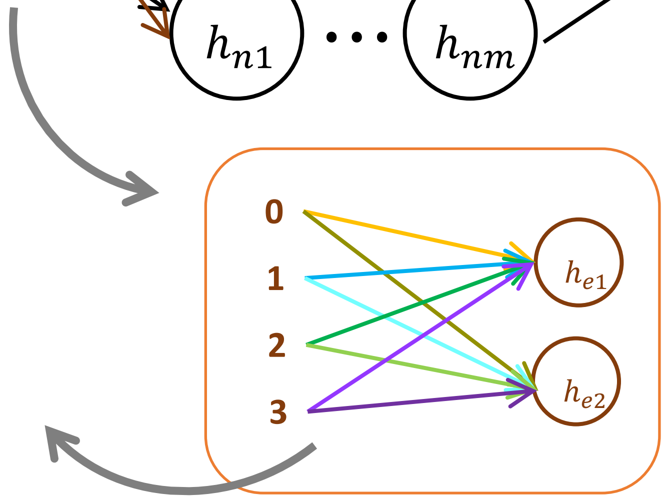

```
neibhd_to_idx={0:'Tender',1:'Civic',2:'FDist',3:'Soma'}
```

Embeddings

- Solution: Embeddings
  - Categories mapped into vectors using embedding weights

Nbhd	rooms	Area(m2)	Price
0	4	61	500k
1	5	72	600k
2	6	83	900k
3	7	91	1200k
…	…	…	…





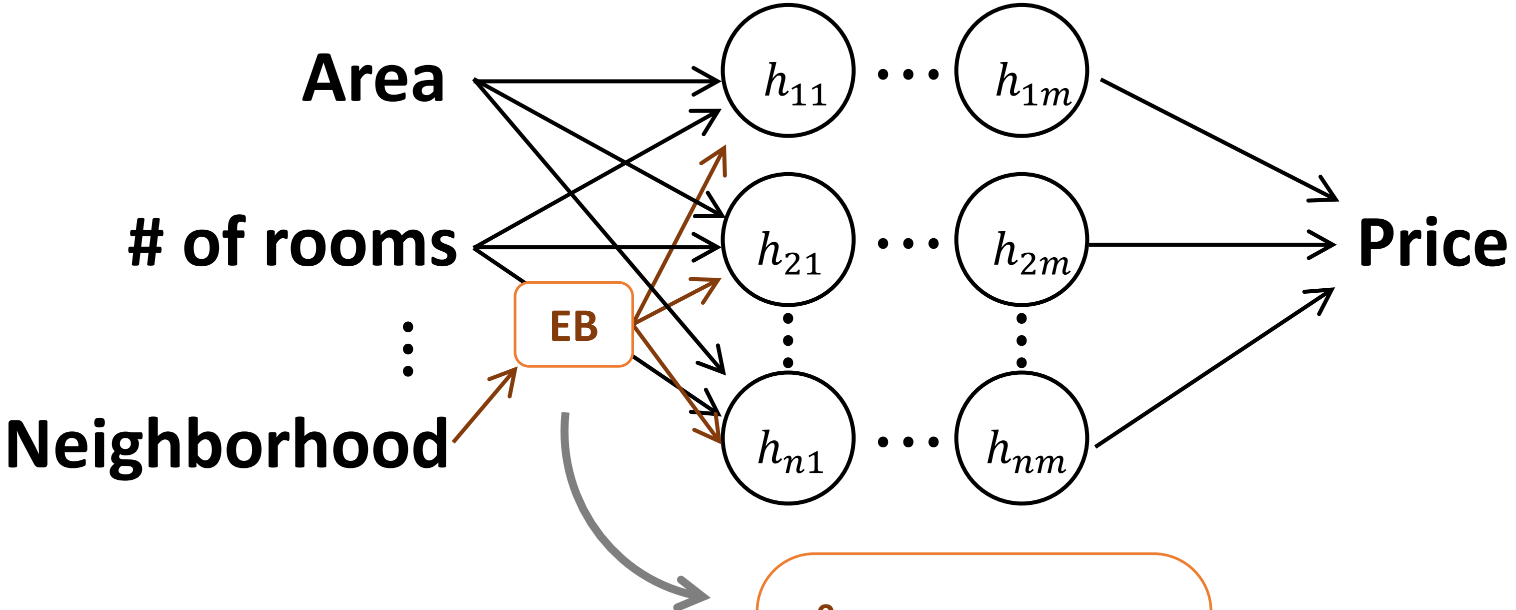
	h_{e1}	h_{e2}
0	w1	w2
1	w3	w4
2	w5	w6
3	w7	w8

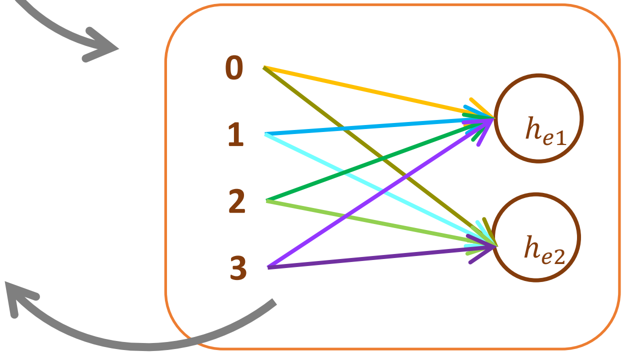
```
neibhd_to_idx={0:'Tender',1:'Civic',2:'FDist',3:'Soma'}
```

Embeddings

- Solution: Embeddings
  - Weights become analog to a lookup table

Nbhd	rooms	Area(m2)	Price
0	4	61	500k
1	5	72	600k
2	6	83	900k
3	7	91	1200k
…	…	…	…





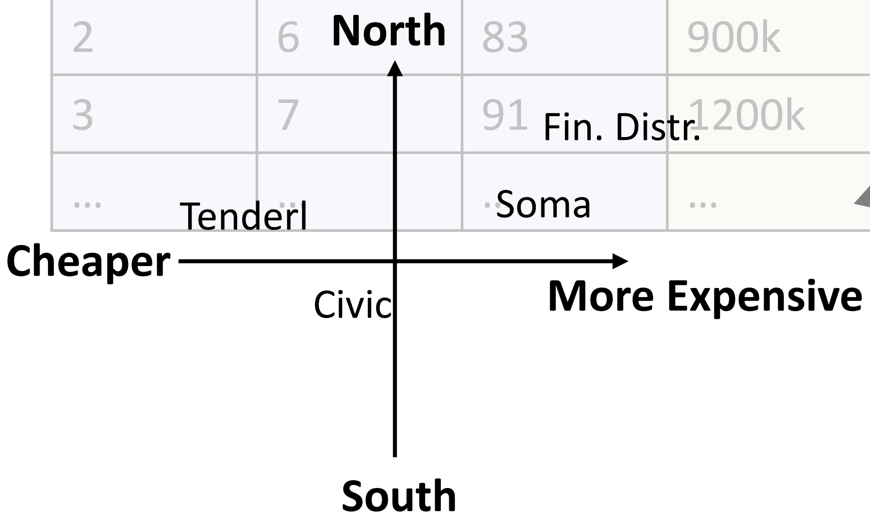
	h_{e1}	h_{e2}
0	w1	w2
1	w3	w4
2	w5	w6
3	w7	w8

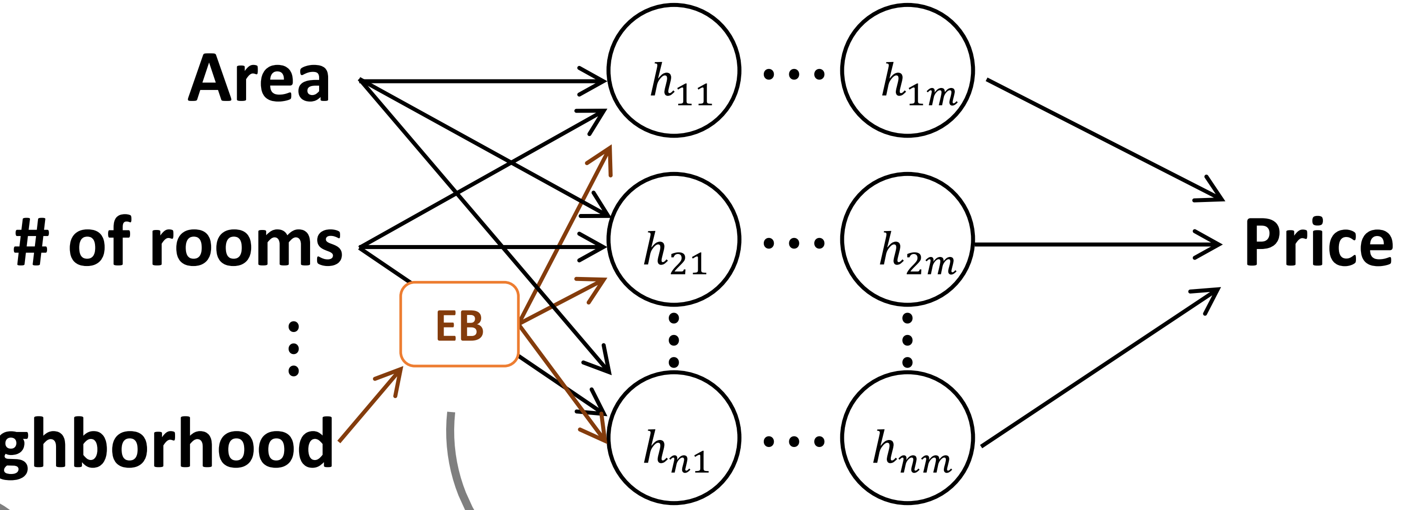
```
neibhd_to_idx={0:'Tender',1:'Civic',2:'FDist',3:'Soma'}
```

Embeddings

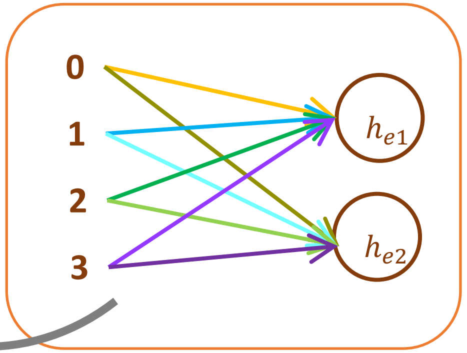
- Solution: Embeddings
  - Weights become analog to a lookup table

Nbhd	rooms	Area(m2)	Price
0	4	61	500k
1	5	72	600k
2	6	83	900k
3	7	91	1200k
...	...	...	...





	h_{e1}	h_{e2}
0	w1	w2
1	w3	w4
2	w5	w6
3	w7	w8




```
neibhd_to_idx={0:'Tender',1:'Civic',2:'FDist',3:'Soma'}
```

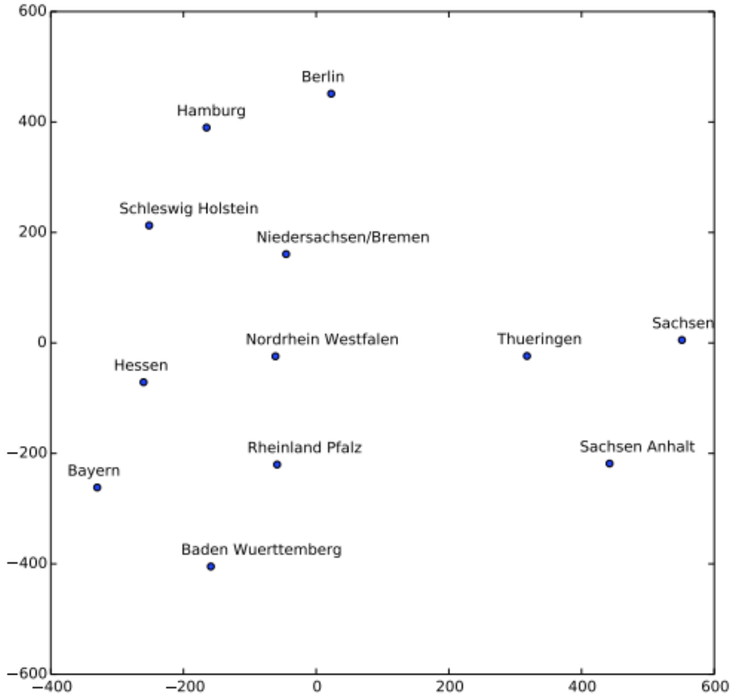
Examples of Popular Embeddings

Rossman Store Sales

- kaggle.com/c/rossmann-store-sales
- **Input:** 8 columns + engineered ones (ex: Store, State, DayOfWeek, Date, Customers, Open, Promo, StateHoliday, SchoolHoliday)
- **Output:** Sales

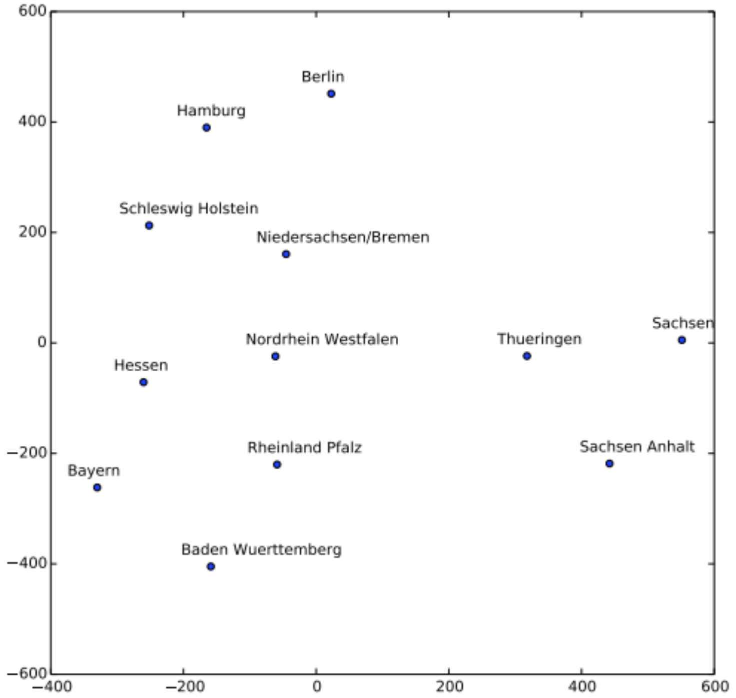
Rossman Store Sales

- kaggle.com/c/rossmann-store-sales
- **Input:** 8 columns + engineered ones (ex: Store, State, DayOfWeek, Date, Customers, Open, Promo, StateHoliday, SchoolHoliday)
- **Output:** Sales
- **Embedding of State (t-SNE projection)**



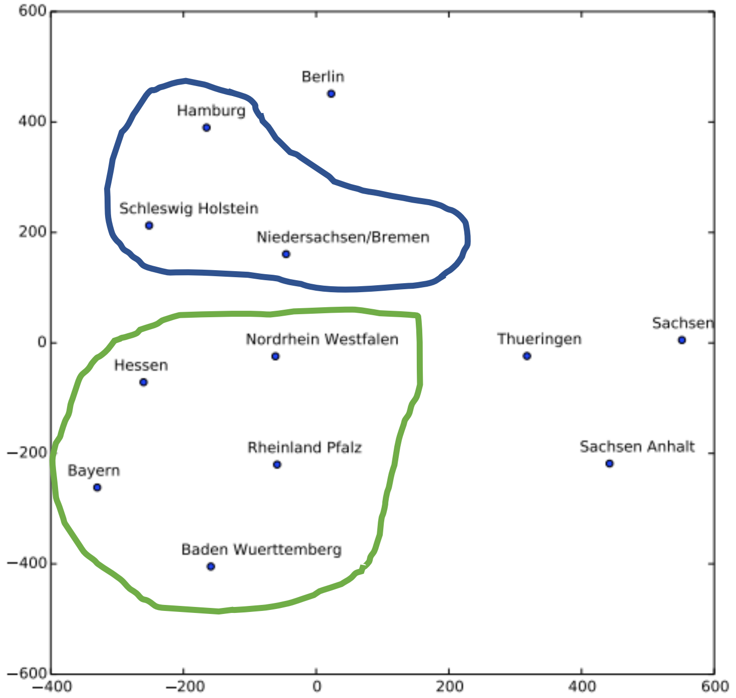
Rossman Store Sales

- kaggle.com/c/rossmann-store-sales
- **Input:** 8 columns + engineered ones (ex: Store, State, DayOfWeek, Date, Customers, Open, Promo, StateHoliday, SchoolHoliday)
- **Output:** Sales
- **Embedding of State**



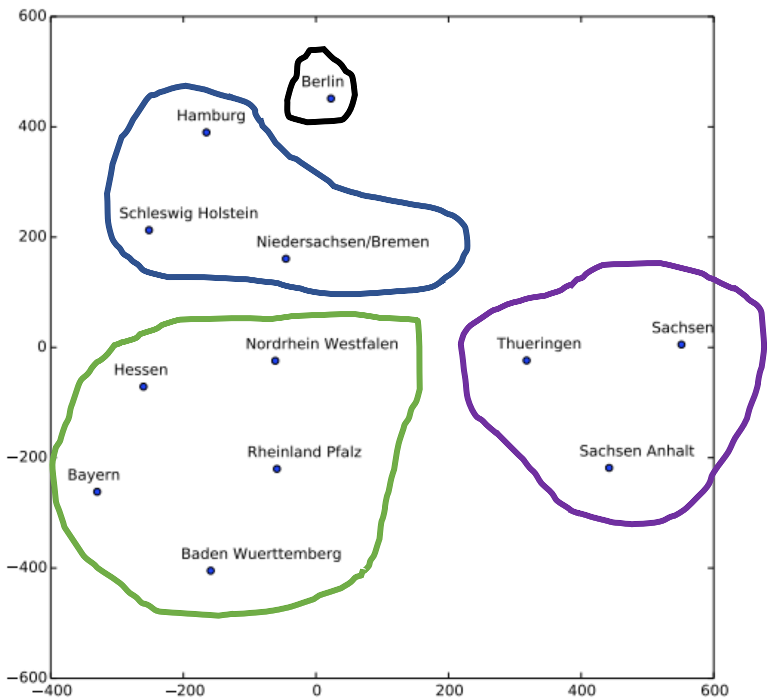
Rossman Store Sales

- kaggle.com/c/rossmann-store-sales
- **Input:** 8 columns + engineered ones (ex: Store, State, DayOfWeek, Date, Customers, Open, Promo, StateHoliday, SchoolHoliday)
- **Output:** Sales
- **Embedding of State**



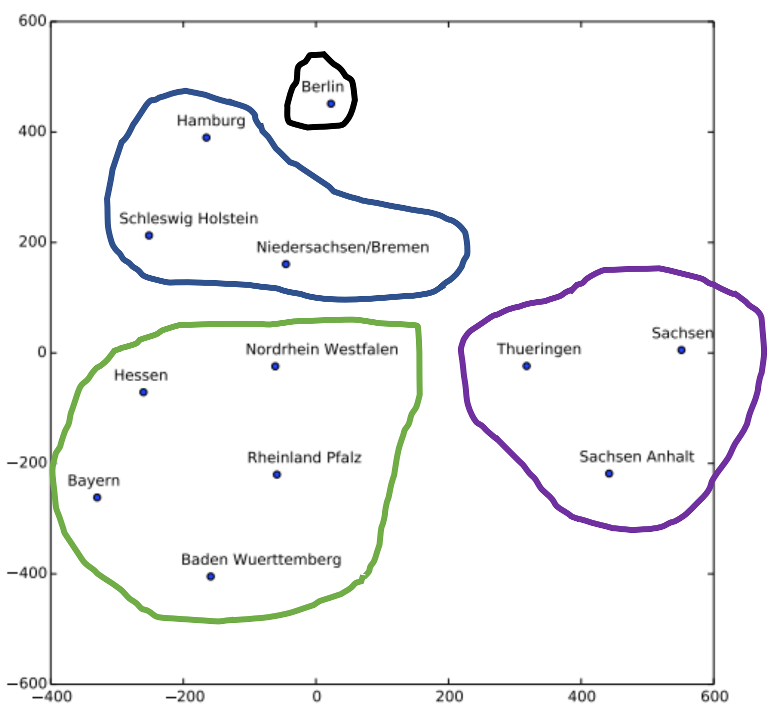
Rossman Store Sales

- kaggle.com/c/rossmann-store-sales
- **Input:** 8 columns + engineered ones (ex: Store, State, DayOfWeek, Date, Customers, Open, Promo, StateHoliday, SchoolHoliday)
- **Output:** Sales
- **Embedding of State**



Rossman Store Sales

- kaggle.com/c/rossmann-store-sales
- **Input:** 8 columns + engineered ones (ex: Store, State, DayOfWeek, Date, Customers, Open, Promo, StateHoliday, SchoolHoliday)
- **Output:** Sales
- **Embedding of State:** Predicts the map of Germany!



Word2Vec

- **Input:** Wikipedia corpus
- **Dimension:** 200
- **Output (CBoW):** The model predicts the current word from a window of surrounding context words
- **Output (CSG):** The model uses the current word to predict the surrounding window of context words

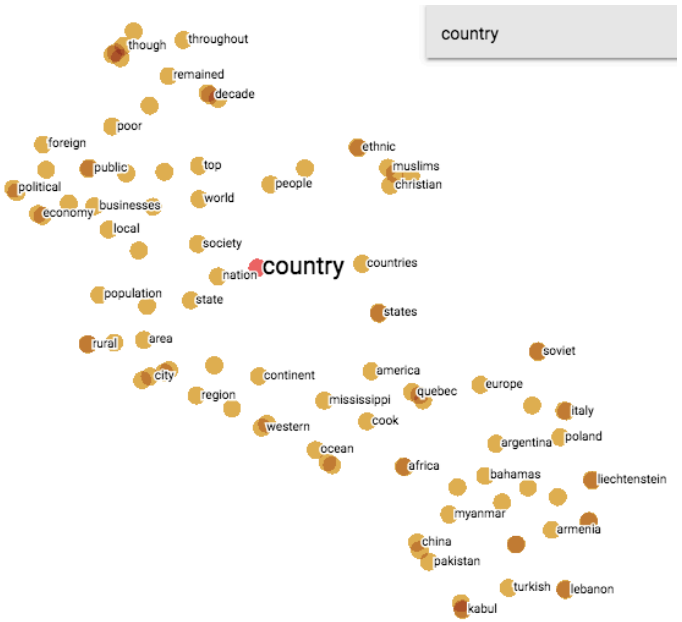
Word2Vec

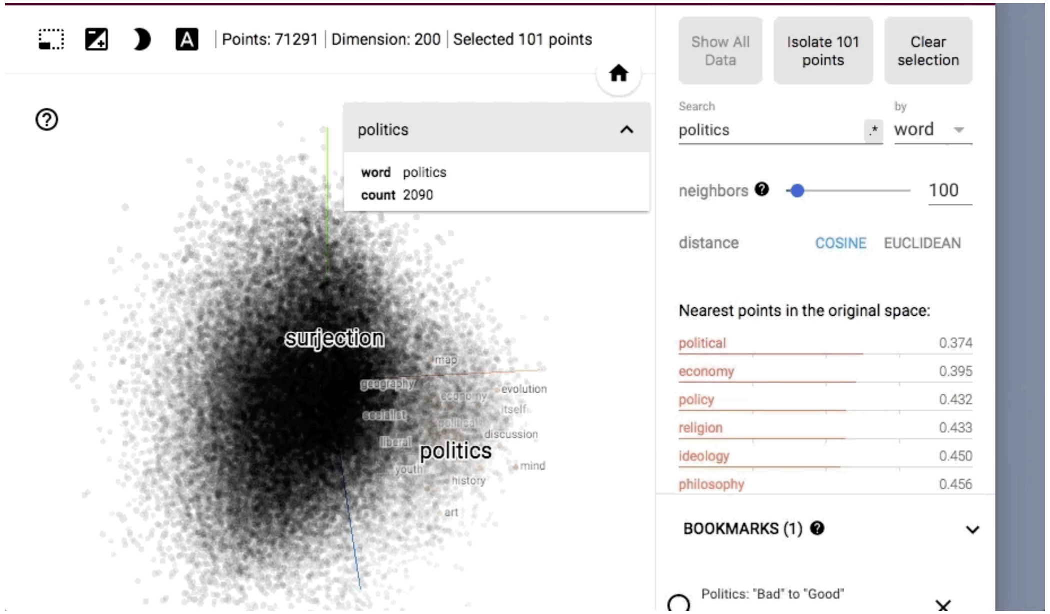
- **Input:** Wikipedia corpus
- **Dimension:** 200
- **Output (CBoW):** The model predicts the current word from a window of surrounding context words
- **Output (CSG):** The model uses the current word to predict the surrounding window of context words

Examples from: http://projector.tensorflow.org/

Word2Vec

- **Input:** Wikipedia corpus
- **Dimension:** 200
- **Output (CBoW):** The model predicts the current word from a window of surrounding context words
- **Output (CSG):** The model uses the current word to predict the surrounding window of context words





Examples from: http://projector.tensorflow.org/

BERT

- **Input:** Wikipedia corpus
- **Output**: Masked words

BERT

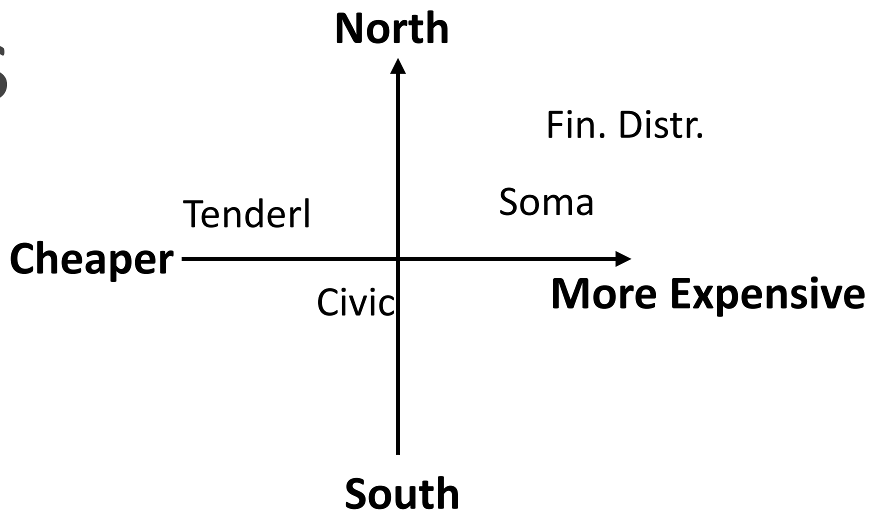
- **Input:** Wikipedia + BookCorpus
- **Output**: Masked words

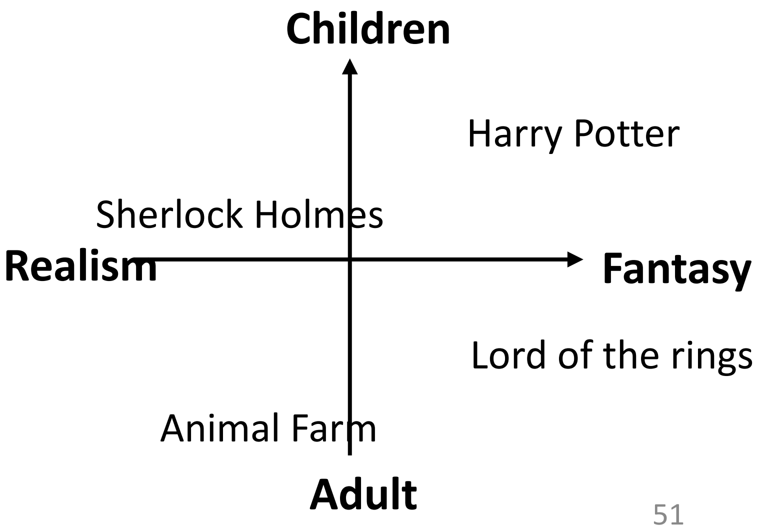
BERT uses a simple approach for this: We mask out 15% of the words in the input, run the entire sequence through a deep bidirectional Transformer encoder, and then predict only the masked words. For example:

```
Input: the man went to the [MASK1] . he bought a [MASK2] of milk.
Labels: [MASK1] = store; [MASK2] = gallon
```

Applications

- Encoding Categorical Variables
- Recommendation Systems
- Mapping words into vectors





Summary

- We use embedding to map categorical variables into a vector
- It overcomes the limitations of one-hot encoding
- The embedding layer is just a hidden layer
- The weights of the embedding serves as a lookup table

Implementing a simplistic example

Next Video