DEPENDENCY PARSING WITH FEATURE ENGINGEERING

Mathijs Mul
Daan van Stigt
May Lee
Selina Blijleven

December 2016

University of Amsterdam

He has good control.

TRANSITION-BASED PARSING

He has good control.

transition	stack	buffer	A
	[ROOT]	[He has good control .]	∅
SHIFT	[ROOT He]	[has good control .]	
SHIFT	[ROOT He has]	[good control .]	
LEFT-REDUCE(nsubj)	[ROOT has]	[good control .]	A∪ nsubj(has, He)
SHIFT	[ROOT has good]	[control .]	
SHIFT	[ROOT has good control]	[.]	
LEFT-REDUCE	[ROOT has control]	[.]	A∪ amod(control, good)
. . .	. . .	. . .	. . .

TRANSITION-BASED PARSING

He has good control.

transition	stack	buffer	A
	[ROOT]	[He has good control .]	∅
SHIFT	[ROOT He]	[has good control .]	
SHIFT	[ROOT He has]	[good control .]	
LEFT-REDUCE(nsubj)	[ROOT has]	[good control .]	A∪ nsubj(has, He)
SHIFT	[ROOT has good]	[control .]	
SHIFT	[ROOT has good control]	[.]	
LEFT-REDUCE	[ROOT has control]	[.]	A∪ amod(control, good)
. . .	. . .	. . .	. . .

TRANSITION-BASED PARSING

He has good control.

transition	stack	buffer	A
	[ROOT]	[He has good control .]	∅
SHIFT	[ROOT He]	[has good control .]	
SHIFT	[ROOT He has]	[good control .]	
LEFT-REDUCE(nsubj)	[ROOT has]	[good control .]	A∪ nsubj(has, He)
SHIFT	[ROOT has good]	[control .]	
SHIFT	[ROOT has good control]	[.]	
LEFT-REDUCE	[ROOT has control]	[.]	A∪ amod(control, good)
. . .	. . .	. . .	. . .

TRANSITION-BASED PARSING

He has good control.

transition	stack	buffer	A
	[ROOT]	[He has good control .]	∅
SHIFT	[ROOT He]	[has good control .]	
SHIFT	[ROOT He has]	[good control .]	
LEFT-REDUCE(nsubj)	[ROOT has]	[good control .]	A∪ nsubj(has, He)
SHIFT	[ROOT has good]	[control .]	
SHIFT	[ROOT has good control]	[.]	
LEFT-REDUCE	[ROOT has control]	[.]	A∪ amod(control, good)
. . .	. . .	. . .	. . .

TRANSITION-BASED PARSING

He has good control.

transition	stack	buffer	A
	[ROOT]	[He has good control .]	∅
SHIFT	[ROOT He]	[has good control .]	
SHIFT	[ROOT He has]	[good control .]	
LEFT-REDUCE(nsubj)	[ROOT has]	[good control .]	A∪ nsubj(has, He)
SHIFT	[ROOT has good]	[control .]	
SHIFT	[ROOT has good control]	[.]	
LEFT-REDUCE	[ROOT has control]	[.]	A∪ amod(control, good)
. . .	. . .	. . .	. . .

TRANSITION-BASED PARSING

He has good control.

transition	stack	buffer	A
	[ROOT]	[He has good control .]	∅
SHIFT	[ROOT He]	[has good control .]	
SHIFT	[ROOT He has]	[good control .]	
LEFT-REDUCE(nsubj)	[ROOT has]	[good control .]	A∪ nsubj(has, He)
SHIFT	[ROOT has good]	[control .]	
SHIFT	[ROOT has good control]	[.]	
LEFT-REDUCE	[ROOT has control]	[.]	A∪ amod(control, good)
. . .	. . .	. . .	. . .

TRANSITION-BASED PARSING

He has good control.

transition	stack	buffer	A
	[ROOT]	[He has good control .]	∅
SHIFT	[ROOT He]	[has good control .]	
SHIFT	[ROOT He has]	[good control .]	
LEFT-REDUCE(nsubj)	[ROOT has]	[good control .]	A∪ nsubj(has, He)
SHIFT	[ROOT has good]	[control .]	
SHIFT	[ROOT has good control]	[.]	
LEFT-REDUCE	[ROOT has control]	[.]	A∪ amod(control, good)
. . .	. . .	. . .	. . .

TRANSITION-BASED PARSING

He has good control.

transition	stack	buffer	A
	[ROOT]	[He has good control .]	∅
SHIFT	[ROOT He]	[has good control .]	
SHIFT	[ROOT He has]	[good control .]	
LEFT-REDUCE(nsubj)	[ROOT has]	[good control .]	A∪ nsubj(has, He)
SHIFT	[ROOT has good]	[control .]	
SHIFT	[ROOT has good control]	[.]	
LEFT-REDUCE	[ROOT has control]	[.]	A∪ amod(control, good)
. . .	. . .	. . .	. . .

TRANSITION-BASED PARSING

He has good control.

transition	stack	buffer	A
	[ROOT]	[He has good control .]	∅
SHIFT	[ROOT He]	[has good control .]	
SHIFT	[ROOT He has]	[good control .]	
LEFT-REDUCE(nsubj)	[ROOT has]	[good control .]	A∪ nsubj(has, He)
SHIFT	[ROOT has good]	[control .]	
SHIFT	[ROOT has good control]	[.]	
LEFT-REDUCE	[ROOT has control]	[.]	A∪ amod(control, good)
. . .	. . .	. . .	. . .

TRANSITION-BASED PARSING

He has good control.

transition	stack	buffer	A
	[ROOT]	[He has good control .]	∅
SHIFT	[ROOT He]	[has good control .]	
SHIFT	[ROOT He has]	[good control .]	
LEFT-REDUCE(nsubj)	[ROOT has]	[good control .]	A∪ nsubj(has, He)
SHIFT	[ROOT has good]	[control .]	
SHIFT	[ROOT has good control]	[.]	
LEFT-REDUCE	[ROOT has control]	[.]	A∪ amod(control, good)
. . .	. . .	. . .	. . .

TRANSITION-BASED PARSING

He has good control.

transition	stack	buffer	A
	[ROOT]	[He has good control .]	∅
SHIFT	[ROOT He]	[has good control .]	
SHIFT	[ROOT He has]	[good control .]	
LEFT-REDUCE(nsubj)	[ROOT has]	[good control .]	A∪ nsubj(has, He)
SHIFT	[ROOT has good]	[control .]	
SHIFT	[ROOT has good control]	[.]	
LEFT-REDUCE	[ROOT has control]	[.]	A∪ amod(control, good)
. . .	. . .	. . .	. . .

TRANSITION-BASED PARSING

He has good control.

transition	stack	buffer	A
	[ROOT]	[He has good control .]	∅
SHIFT	[ROOT He]	[has good control .]	
SHIFT	[ROOT He has]	[good control .]	
LEFT-REDUCE(nsubj)	[ROOT has]	[good control .]	A∪ nsubj(has, He)
SHIFT	[ROOT has good]	[control .]	
SHIFT	[ROOT has good control]	[.]	
LEFT-REDUCE	[ROOT has control]	[.]	A∪ amod(control, good)
. . .	. . .	. . .	. . .

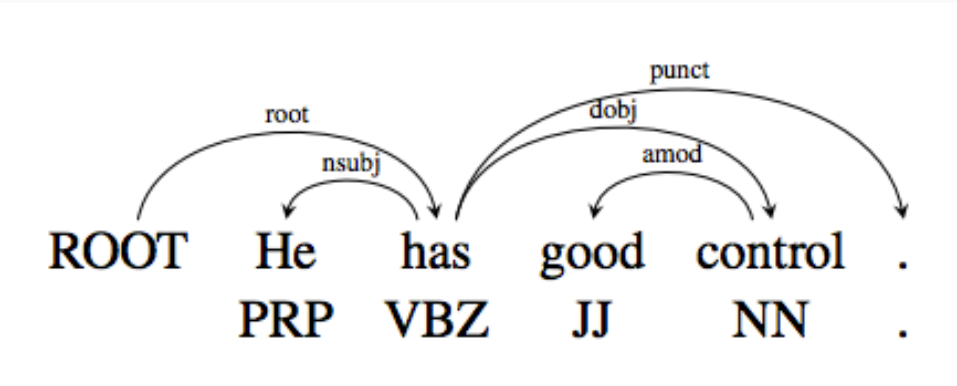
TRANSITION-BASED PARSING

He has good control.

transition	stack	buffer	A
	[ROOT]	[He has good control .]	∅
SHIFT	[ROOT He]	[has good control .]	
SHIFT	[ROOT He has]	[good control .]	
LEFT-REDUCE(nsubj)	[ROOT has]	[good control .]	A∪ nsubj(has, He)
SHIFT	[ROOT has good]	[control .]	
SHIFT	[ROOT has good control]	[.]	
LEFT-REDUCE	[ROOT has control]	[.]	A∪ amod(control, good)
. . .	. . .	. . .	. . .

He has good control.

transition	stack	buffer	A
	[ROOT]	[He has good control .]	∅
SHIFT	[ROOT He]	[has good control .]	
SHIFT	[ROOT He has]	[good control .]	
LEFT-REDUCE(nsubj)	[ROOT has]	[good control .]	A∪ nsubj(has, He)
SHIFT	[ROOT has good]	[control .]	
SHIFT	[ROOT has good control]	[.]	
LEFT-REDUCE	[ROOT has control]	[.]	A∪ amod(control, good)
. . .	. . .	. . .	. . .



INFER TRANSITIONS FROM DATA

Infer configurations and transitions from training data:

```
1       He      _       PRP     PRP     _       2       nsubj
2       has     _       VBZ     VBZ     _       0       root
3       good    _       JJ      JJ      _       4       amod
4       control _       NN      NN      _       2       dobj
5       .       _       .       .       _       2       punct
```

INFER TRANSITIONS FROM DATA

Infer configurations and transitions from training data:

```
1       He      _       PRP     PRP     _       2       nsubj
2       has     _       VBZ     VBZ     _       0       root
3       good    _       JJ      JJ      _       4       amod
4       control _       NN      NN      _       2       dobj
5       .       _       .       .       _       2       punct
```

transition	stack	buffer	A
	[ROOT]	[He has good control .]	∅
SHIFT	[ROOT He]	[has good control .]	
SHIFT	[ROOT He has]	[good control .]	
LEFT-REDUCE(nsubj)	[ROOT has]	[good control .]	A∪ nsubj(has, He)
SHIFT	[ROOT has good]	[control .]	
SHIFT	[ROOT has good control]	[.]	
LEFT-REDUCE	[ROOT has control]	[.]	A∪ amod(control, good)
. . .	. . .	. . .	. . .

INFER TRANSITIONS FROM DATA

Infer configurations and transitions from training data:

```
1       He      _       PRP     PRP     _       2       nsubj
2       has     _       VBZ     VBZ     _       0       root
3       good    _       JJ      JJ      _       4       amod
4       control _       NN      NN      _       2       dobj
5       .       _       .       .       _       2       punct
```

transition	stack	buffer	A
	[ROOT]	[He has good control .]	∅
SHIFT	[ROOT He]	[has good control .]	
SHIFT	[ROOT He has]	[good control .]	
LEFT-REDUCE(nsubj)	[ROOT has]	[good control .]	A∪ nsubj(has, He)
SHIFT	[ROOT has good]	[control .]	
SHIFT	[ROOT has good control]	[.]	
LEFT-REDUCE	[ROOT has control]	[.]	A∪ amod(control, good)
· · ·	· · ·	· · ·	· · ·

INFER TRANSITIONS FROM DATA

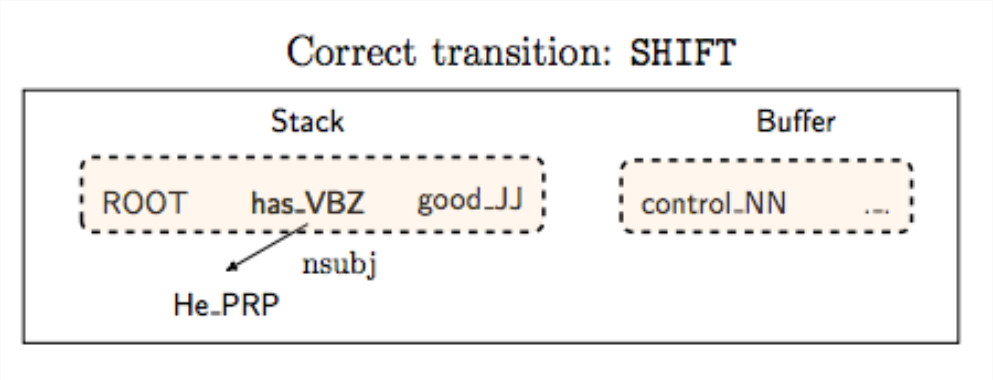
Infer configurations and transitions from training data:

```
1       He      _       PRP     PRP     _       2       nsubj
2       has     _       VBZ     VBZ     _       0       root
3       good    _       JJ      JJ      _       4       amod
4       control _       NN      NN      _       2       dobj
5       .       _       .       .       _       2       punct
```

transition	stack	buffer	A
	[ROOT]	[He has good control .]	∅
SHIFT	[ROOT He]	[has good control .]	
SHIFT	[ROOT He has]	[good control .]	
LEFT-REDUCE(nsubj)	[ROOT has]	[good control .]	A∪ nsubj(has, He)
SHIFT	[ROOT has good]	[control .]	
SHIFT	[ROOT has good control]	[.]	
LEFT-REDUCE	[ROOT has control]	[.]	A∪ amod(control, good)
. . .	. . .	. . .	. . .

FEATURE EXTRACTION

- A selection of the 48 features used, following Chen & Manning (2014):
  - words (18)
  - POS tags (18)
  - labels (12)



- In the example above:
  - $s_1 = \text{good}$
  - $b1.t = \text{NN}$
  - $lc_1(s_2).t = \text{PRP}$
  - $lc_1(s_2).l = \text{nsubj}$
  - $rc_1(s_2) = \text{NONE}$

VECTOR REPRESENTATIONS

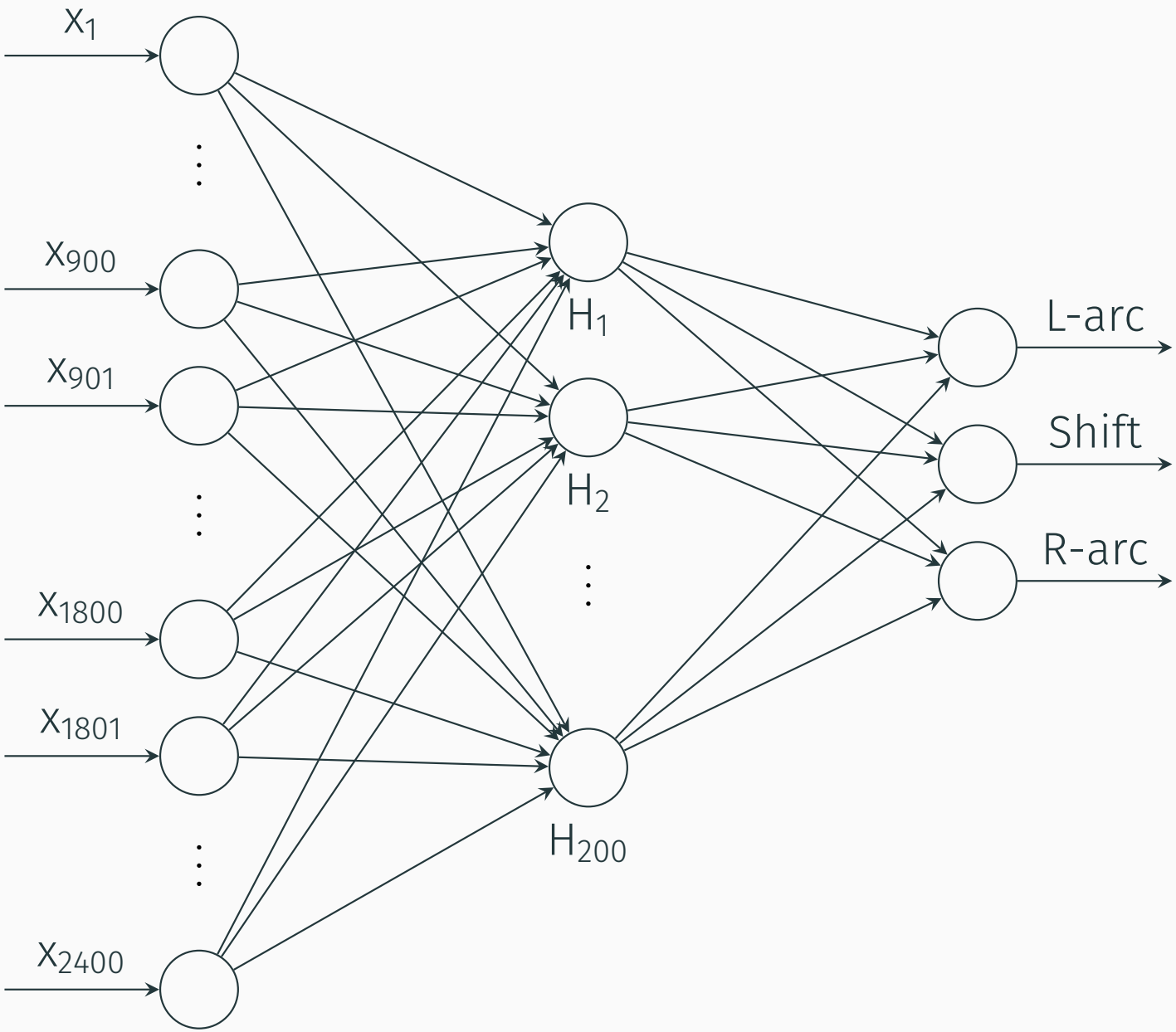
- Each feature is represented as a 50-dimensional vector.

$$s_1 = \text{good} \rightarrow \begin{bmatrix} x_1 \\ \vdots \\ x_{50} \end{bmatrix} \qquad s_2 = \text{has} \rightarrow \begin{bmatrix} x_{51} \\ \vdots \\ x_{100} \end{bmatrix} \qquad \dots$$

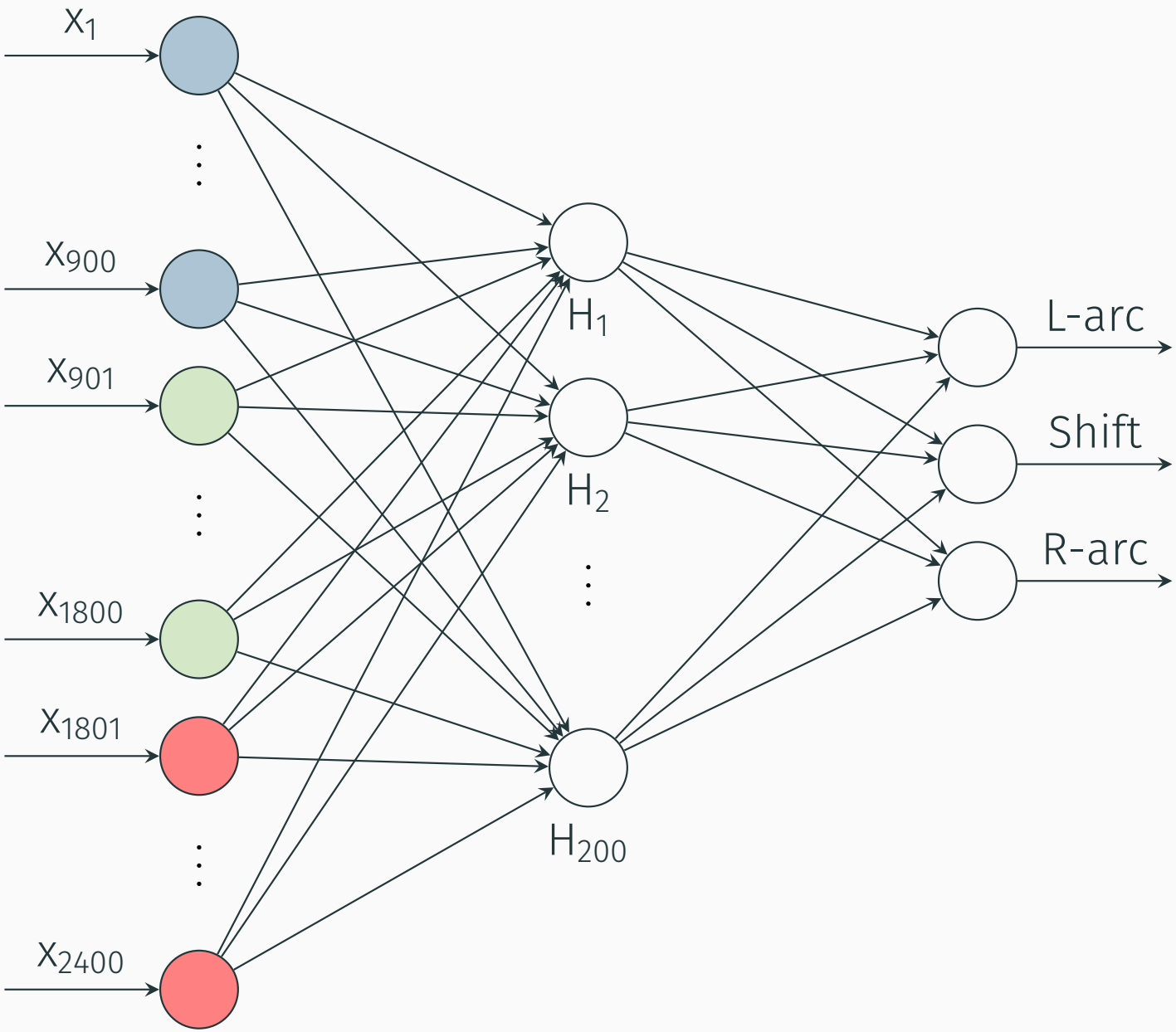
- These are stacked on top of each other to give the full input to the Neural Network: $48 \cdot 50 = 2400$.

$$\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_1.t \\ s_2.t \\ \vdots \\ lc_1(s_1) \\ lc_1(s_2) \\ \vdots \end{bmatrix} \rightarrow \begin{bmatrix} x_1 \\ \vdots \\ x_{900} \\ x_{901} \\ \vdots \\ x_{1800} \\ x_{1801} \\ \vdots \\ x_{2400} \end{bmatrix}$$

TRAINING A NEURAL NETWORK



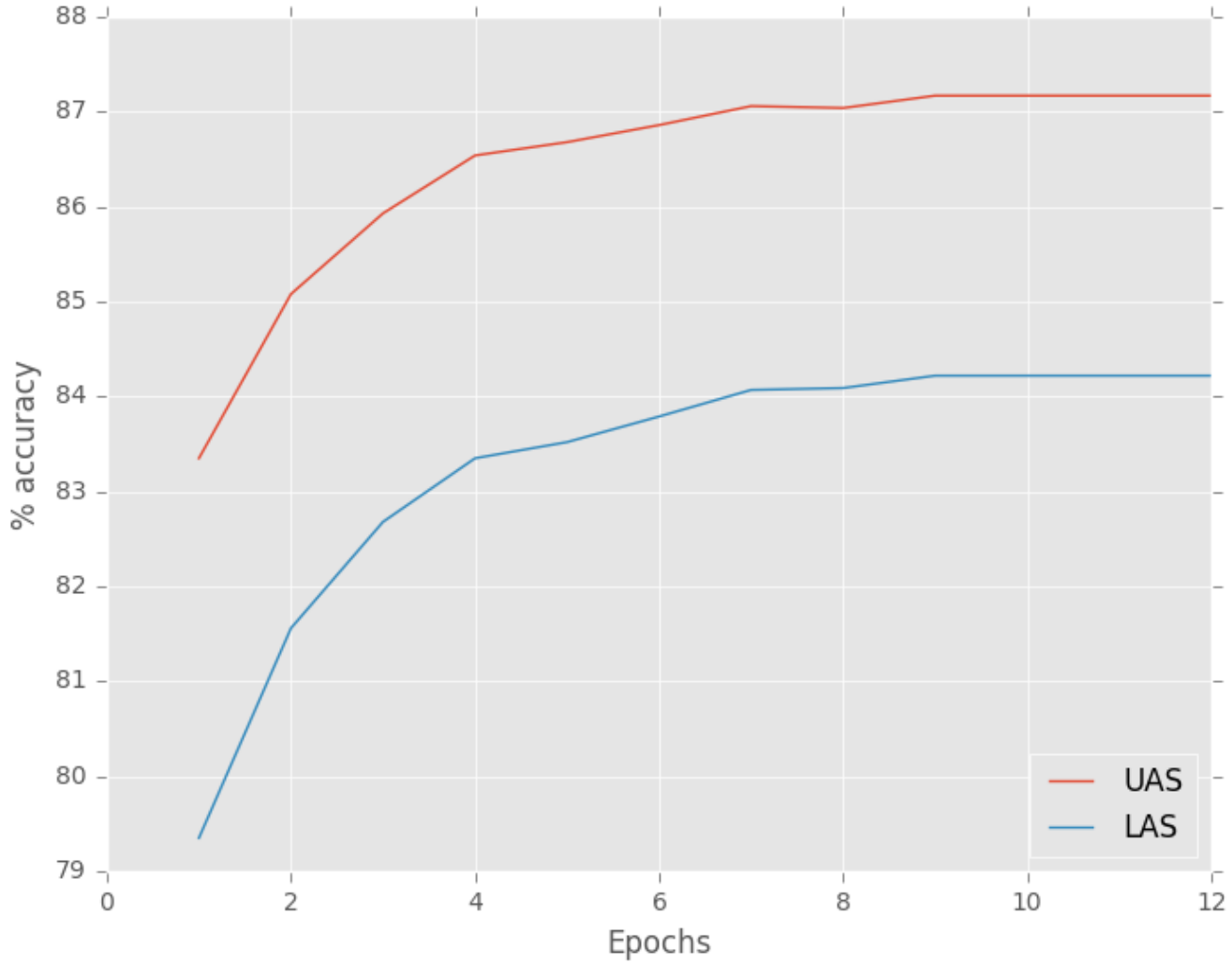
TRAINING A NEURAL NETWORK



BASELINE RESULTS

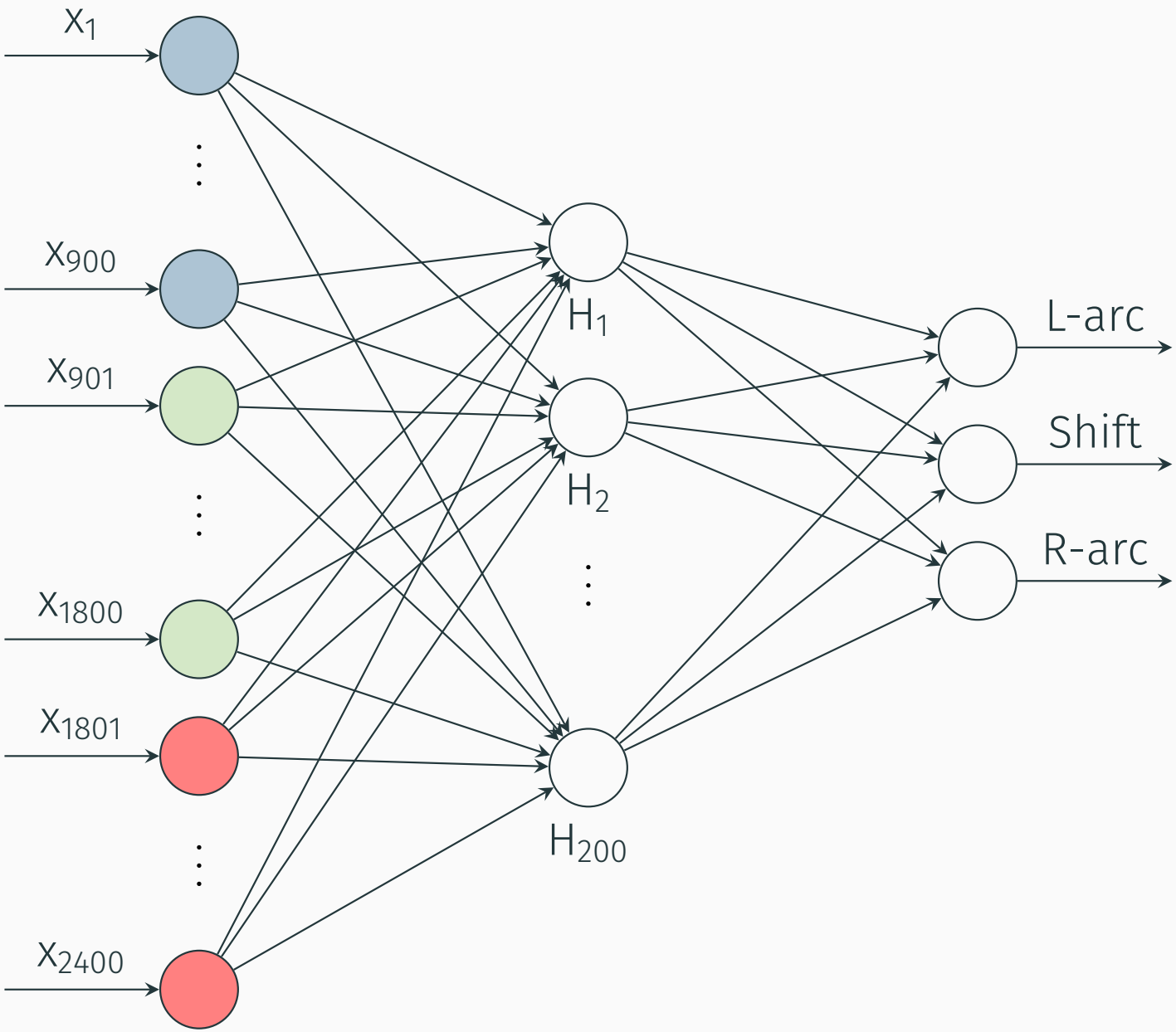
BASELINE RESULTS

Development set (+/- 1800 sentences)

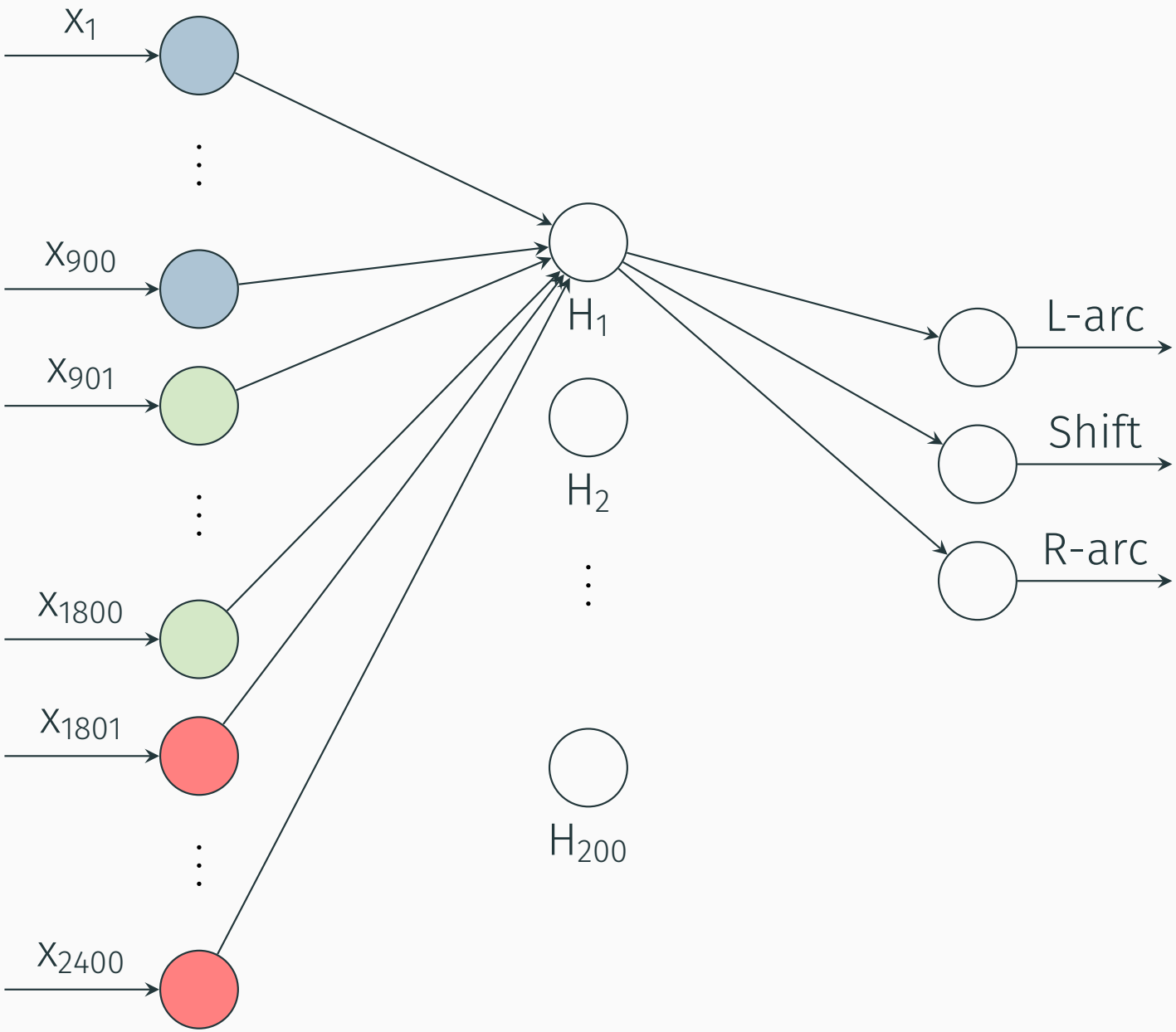


INTERPRETING THE TRAINED WEIGHTS

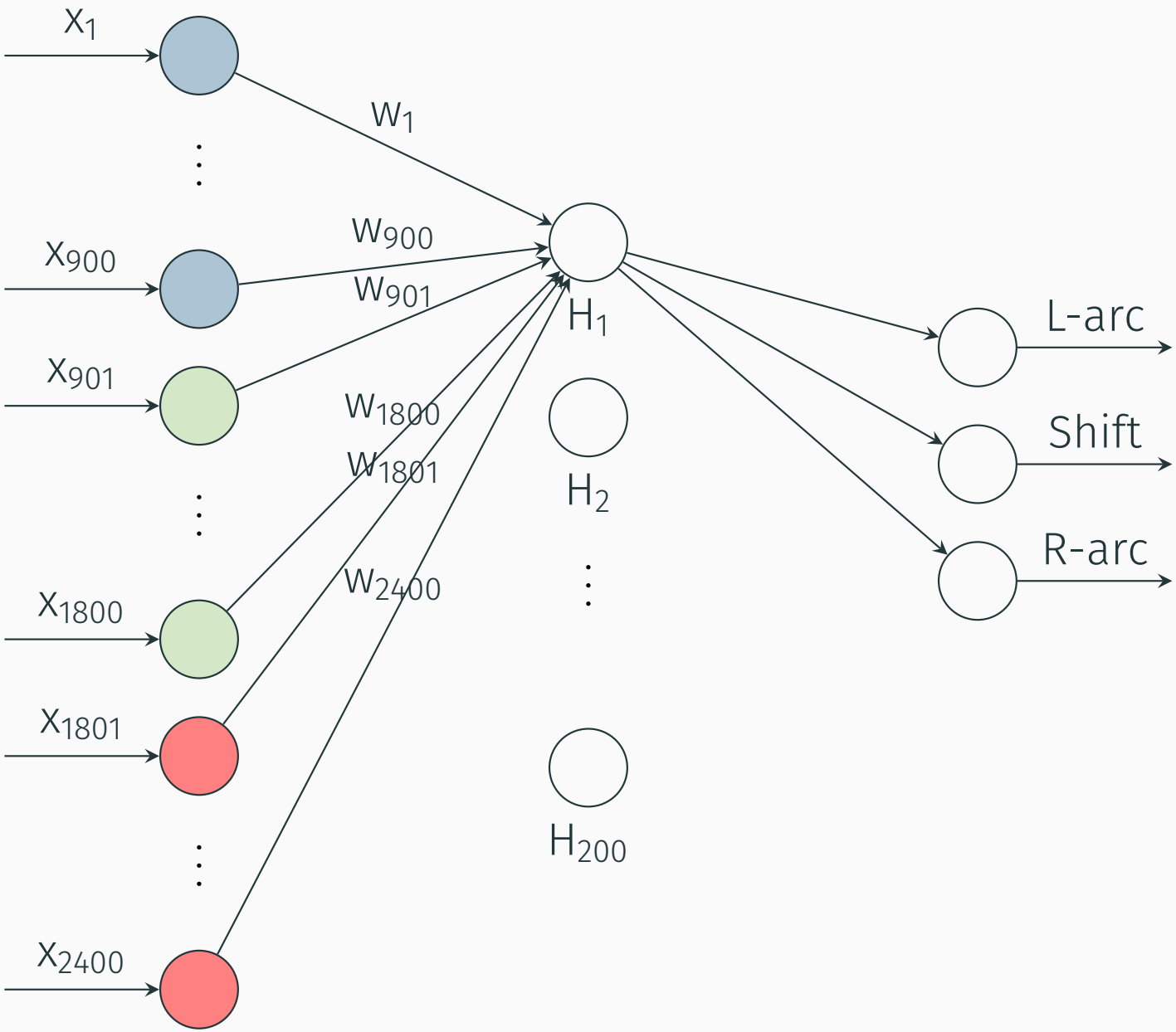
INTERPRETING THE TRAINED WEIGHTS



INTERPRETING THE TRAINED WEIGHTS



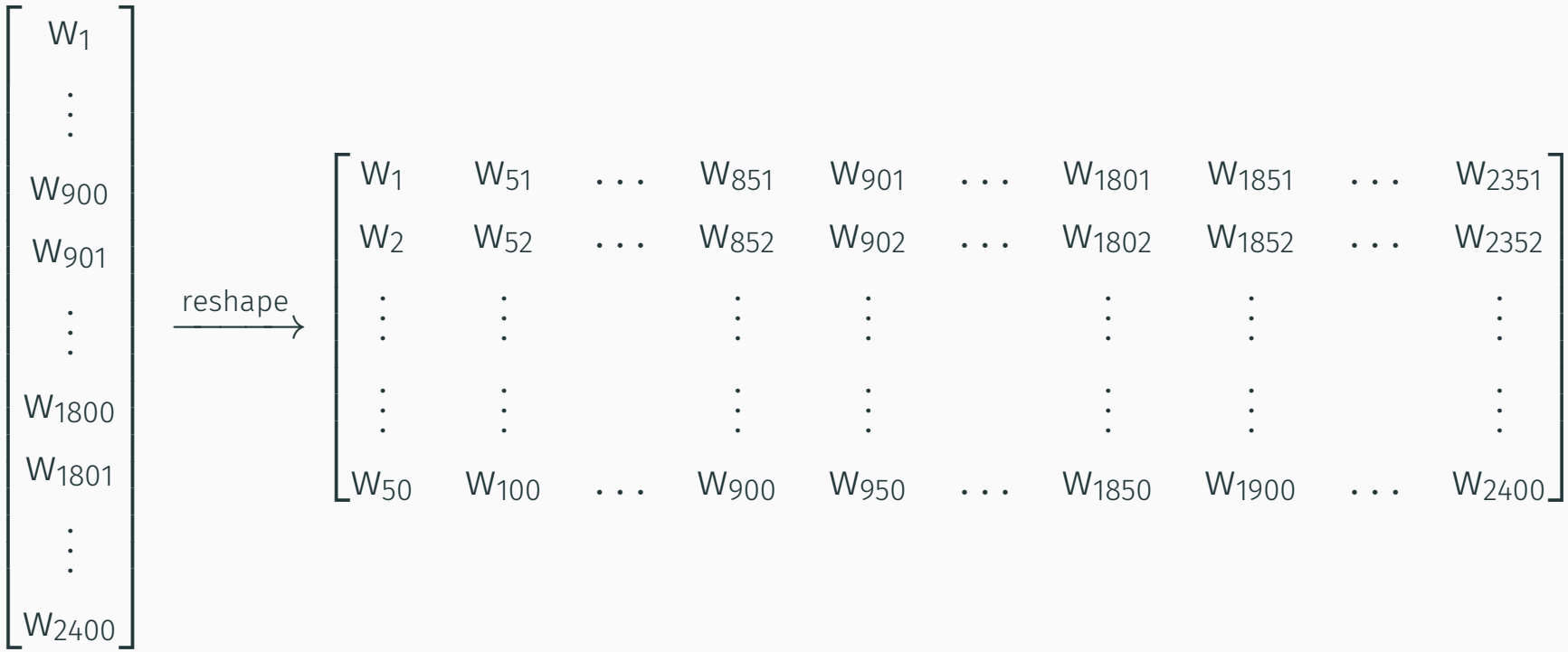
INTERPRETING THE TRAINED WEIGHTS



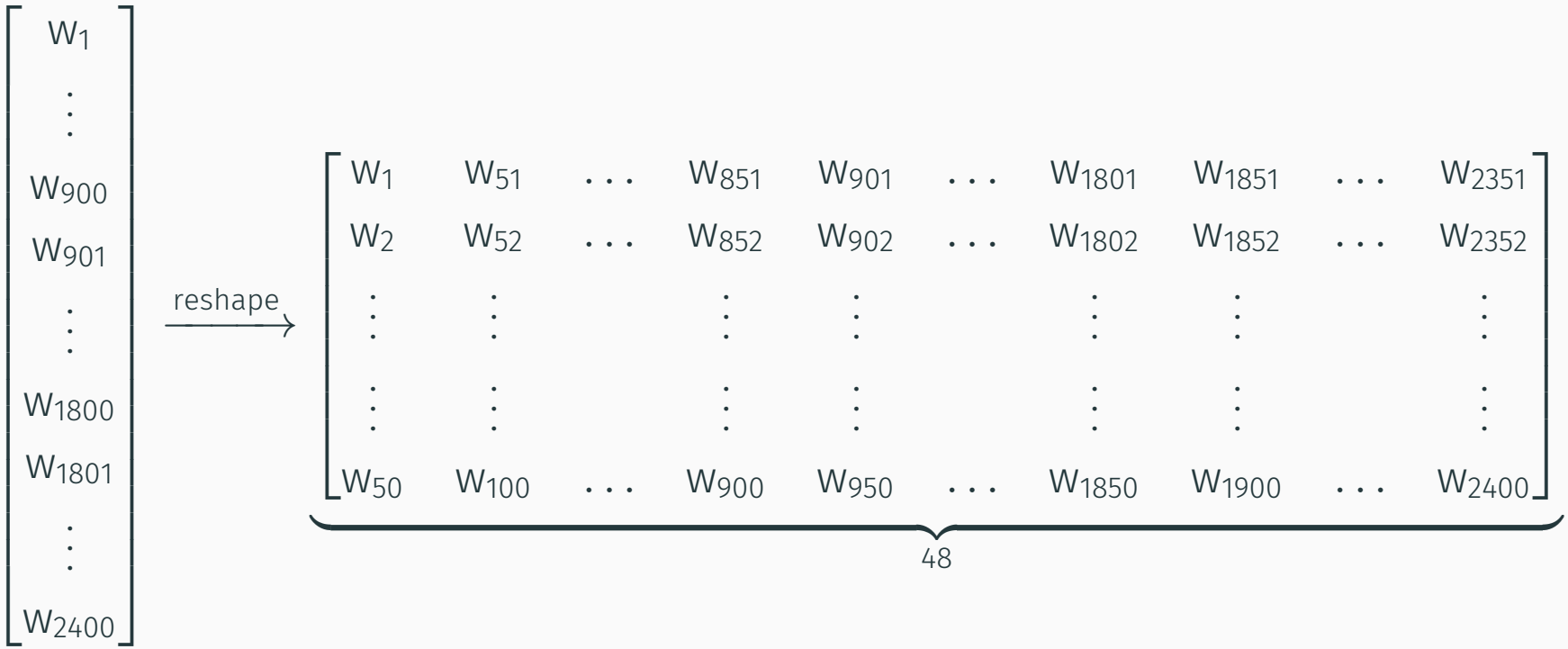
INTERPRETING THE TRAINED WEIGHTS

$$\begin{bmatrix} w_1 \\ \vdots \\ w_{900} \\ w_{901} \\ \vdots \\ w_{1800} \\ w_{1801} \\ \vdots \\ w_{2400} \end{bmatrix}$$

INTERPRETING THE TRAINED WEIGHTS

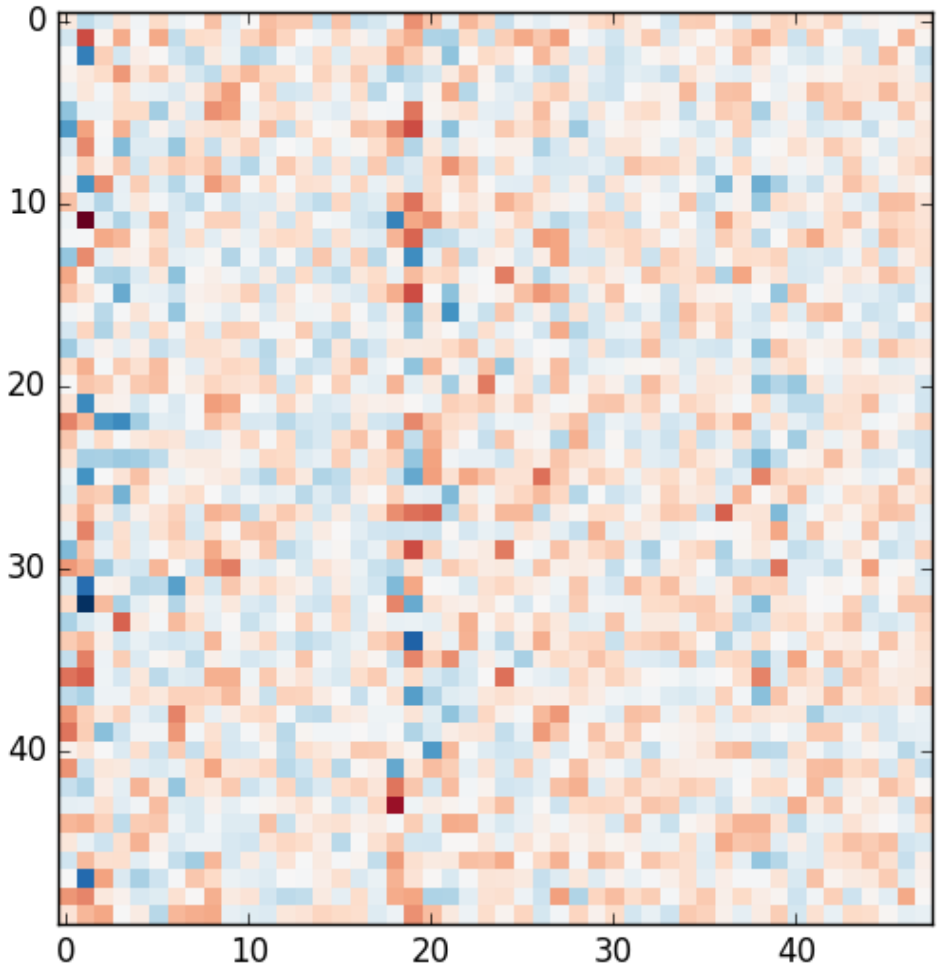


INTERPRETING THE TRAINED WEIGHTS



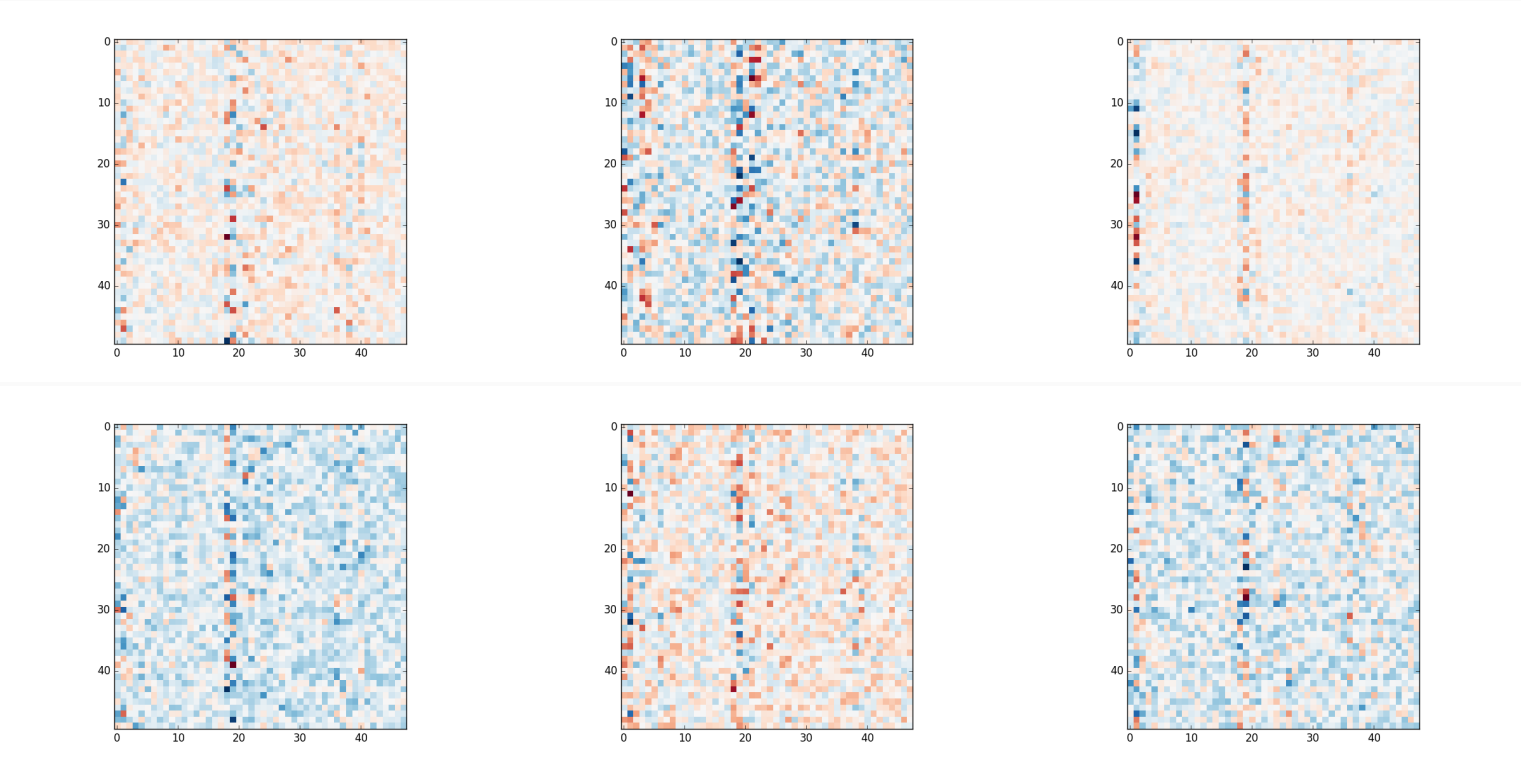
INTERPRETING THE TRAINED WEIGHTS

Heatmap of reshaped weight vector W_1. Blue indicates negative value, red indicates positive value. Lighter shades are closer to zero.



INTERPRETING THE TRAINED WEIGHTS

Heatmaps of reshaped weight vectors W_i for nodes $i = 1, \ldots, 6$. Blue indicates negative value, red indicates positive value. Lighter shades are closer to zero.

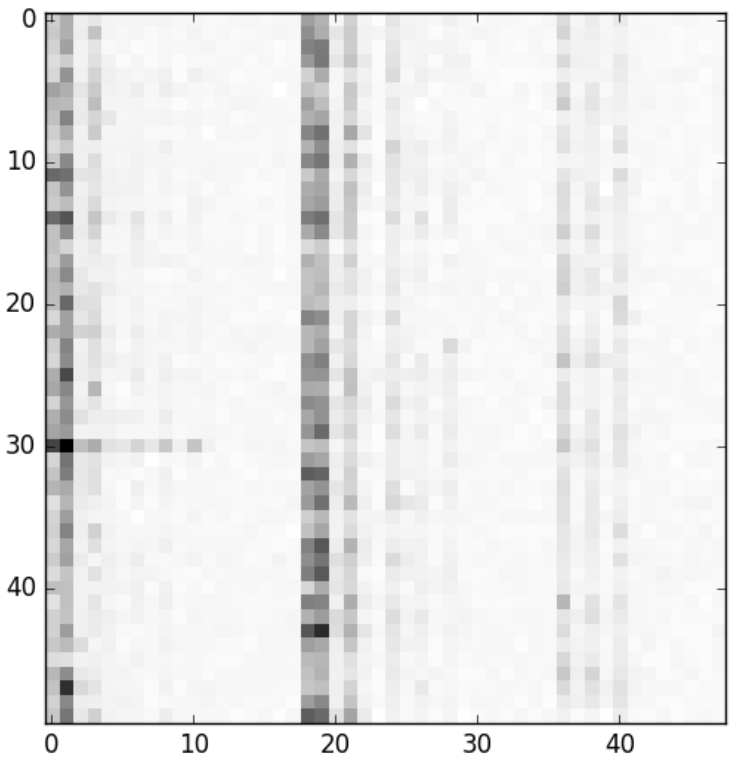


INTERPRETING THE TRAINED WEIGHTS

Heatmap of $\sum_{i=1}^{200} |W_i|$

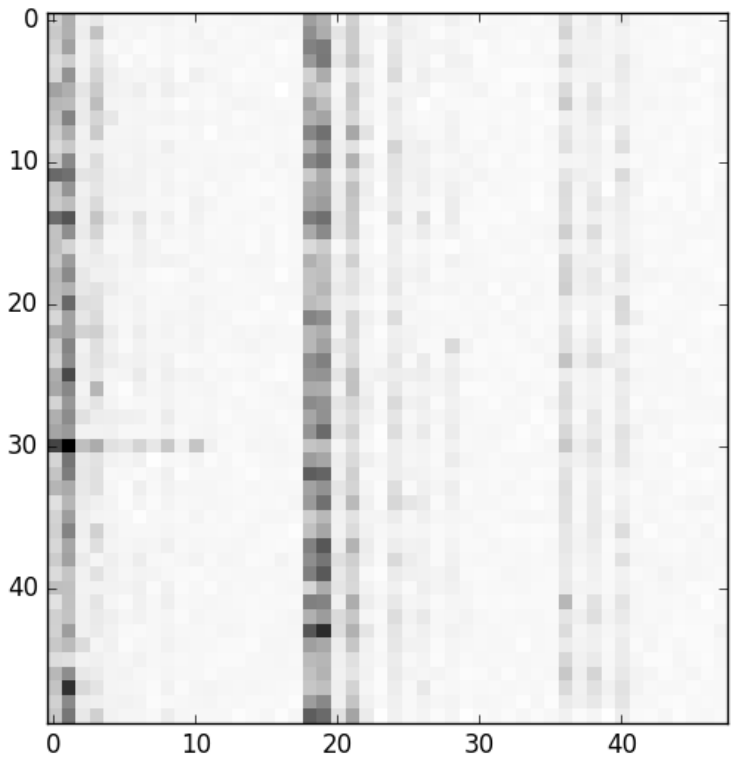
INTERPRETING THE TRAINED WEIGHTS

Heatmap of $\sum_{i=1}^{200} |W_i|$



INTERPRETING THE TRAINED WEIGHTS

Heatmap of $\sum_{i=1}^{200} |W_i|$



Hypothesis: most important features are in columns $0, 1$ and $18, 19$. These are features s_1, s_2 and $s_1.t, s_2.t$.

REMOVING FEATURES

REMOVING FEATURES

$$[0, 1, 18, 19] = [s_1, s_2, s_1.ts_2.t]$$

REMOVING FEATURES

$$[0, 1, 18, 19] = [s_1, s_2, s_1.ts_2.t]$$
$$[10 - 17, 22 - 35, 40 - 47] = [lc_1(lc_1(s_1)), \ldots, lc_1(lc_1(s_1)).t, \ldots]$$

REMOVING FEATURES

$$[0, 1, 18, 19] = [s_1, s_2, s_1.ts_2.t]$$
$$[10-17, 22-35, 40-47] = [lc_1(lc_1(s_1)), \ldots, lc_1(lc_1(s_1)).t, \ldots]$$