

# CS 224N - Assignment 3

Ilya Kachan

July 9, 2021

## 1 Machine Learning & Neural Networks

(a) i. Momentum technique makes gradient updates more smooth since it uses not the gradient vector itself at each step, but the weighted average of all gradients from past steps with exponentially decaying weights. This allows to reduce the gradient variance and makes the optimization more "directional", since the gradient direction doesn't change from step to step as much as in the ordinary gradient decay.

(a) ii. Dividing by  $\sqrt{v}$  means giving more update to the weights, which gradients are close to zero. Potentially this trick could help the optimization procedure to get out of local minimums.

(b) i. In dropout  $\gamma = 1 - p$ , since  $Eh_{drop,i} = h_i(1 - p) + 0 \cdot p = (1 - p)h_i$ .

(b) ii. Applying dropout to a neural network amounts to sampling a thinned network from it. For each presentation of each training case, a new thinned network is sampled and trained. So training a neural network with dropout can be seen as training a collection of  $2^n$  thinned networks with extensive weight sharing, where each thinned network gets trained very rarely, if at all. At test time, it is not feasible to explicitly average the predictions from exponentially many thinned models.

## 2 Neural Transition-Based Dependency Parsing

(a) Let's first parse a sentence "I parsed this sentence correctly" manually:

Stack	Buffer	New Dependency	Transition
[ROOT]	[I, parsed, this, sentence, correctly]		Initial Configuration
[ROOT, I]	[parsed, this, sentence, correctly]		SHIFT
[ROOT, I, parsed]	[this, sentence, correctly]		SHIFT
[ROOT, parsed]	[this, sentence, correctly]	parsed $\rightarrow$ I	LEFT-ARC
[ROOT, parsed, this]	[sentence, correctly]		SHIFT
[ROOT, parsed, this, sentence]	[correctly]		SHIFT
[ROOT, parsed, sentence]	[correctly]	sentence $\rightarrow$ this	LEFT-ARC
[ROOT, parsed]	[correctly]	parsed $\rightarrow$ sentence	RIGHT-ARC
[ROOT, parsed, correctly]	[]		SHIFT
[ROOT, parsed]	[]	parsed $\rightarrow$ correctly	RIGHT-ARC
[ROOT]	[]	ROOT $\rightarrow$ parsed	RIGHT-ARC

(b) Since each parsing step either decreases the stack size by 1 word or decreases the buffer size by 1 word, a sentence with  $n$  words will be parsed with  $2n$  steps.

(e) The best results my model has achieved are: 88.95 UAS on test set and 88.76 UAS on dev set (7th epoch).

(f)

i. *I disembarked and was heading to a wedding fearing my death.*

- **Error type:** Verb phrase attachment error.
- **Incorrect dependency:** wedding  $\rightarrow$  fearing.
- **Correct dependency:** heading  $\rightarrow$  fearing.

ii. *It makes me want to rush out and rescue people from dilemmas of their own making.*

- **Error type:** Coordination attachment error.
- **Incorrect dependency:** makes  $\rightarrow$  rescue.
- **Correct dependency:** rush  $\rightarrow$  rescue.

iii. *It is on loan from a guy named Joe O'Neill in Midland, Texas.*

- **Error type:** Prepositional phrase attachment error.
- **Incorrect dependency:** named  $\rightarrow$  Midland.
- **Correct dependency:** load  $\rightarrow$  Midland.

iv. *Brian has been one of the most crucial elements to the success of Mozilla software.*

- **Error type:** Prepositional phrase attachment error.
- **Incorrect dependency:** elements  $\rightarrow$  most.
- **Correct dependency:** crucial  $\rightarrow$  most.