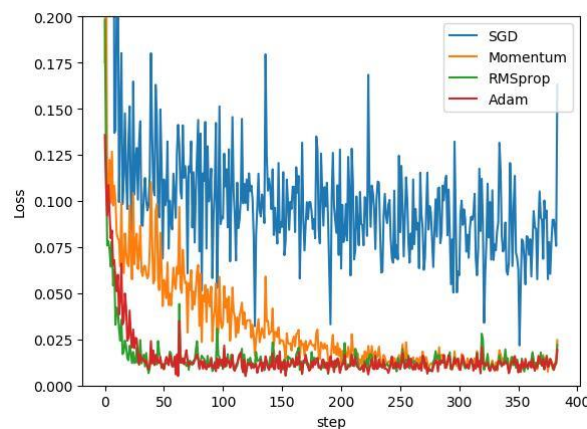


NLP Assignment 3 Sara Kodeiri 96521443

Q1 – Adam Optimizer

i - In stochastic gradient descent, the parameters are immediately updated according to the latest gradient of the loss function. By having a momentum, Adam has created some sort of memory for all the gradients we've encountered up to that point. With beta, we can decide how much the new information can influence the overall momentum. (It's important to notice that if beta is zero, we have the normal SGD). With alpha, we decide how much the momentum influences the parameters. By having the momentum, each result is closer to its previous step because only one new value has been taken into account which in the end, reduces the variance. With less variance, we get to the desired result sooner and don't need to test a lot of the values in between.



ii – According to the formulas, the larger the adaptive learning rate is, the smaller update it will get. When v is small, it means that it's not changing rapidly and might be stuck at a flat point. By making the update more drastic, it helps it get out of that point and make the convergence more efficient.

Q1 – Dropout

i –
$$[p_{\text{drop}} * 0 + (1 - p_{\text{drop}}) * h_i * \text{lambda}] = h_i$$

$$\Rightarrow \text{lambda} = 1 / (1 - p_{\text{drop}})$$

ii – Dropout is applied during training to reduce the chances of the model getting overfitted. Random dropout breaks up these co-adaptations by making the presence of any particular hidden unit unreliable. (Srivastava et al.)

While dropout facilitates learning, using it at evaluation will completely ignore some neurons in the hidden layers which have accurately trained weights. This will result in randomized results in evaluation and won't give us a full-proof prediction.

Q2 – Sentence Parsing

Stack	Buffer	New Dependency	Transition
ROOT	I, Parsed, This, Sentence, Correctly		Init
ROOT, I	Parsed, This, Sentence, Correctly		S
ROOT, I, Parsed	This, Sentence, Correctly		S
ROOT, Parsed	This, Sentence, Correctly	Parsed to I	LA
ROOT, Parsed, This	Sentence, Correctly		S
ROOT, Parsed, This, Sentence	Correctly		S
ROOT, Parsed, Sentence	Correctly	Sentence to This	LA
ROOT, Parsed	Correctly	Parsed to Sentence	RA
ROOT, Parsed, Correctly	<empty>		S
ROOT, Parsed	<empty>	Parsed to Correctly	RA
ROOT	<empty>	ROOT to Parsed	RA

Q2 – Count

Taking the initial configuration into account, we have $2n + 1$ steps. One for the initialization, n for the number of shifts (because each word eventually comes into the stack from the buffer) and n left or right arcs; because each word has an arc transition as well.

Q2 – Errors

i –

Verb Phrase Attachment Error

Incorrect: wedding to fearing

Correct: heading to fearing

ii –

Coordination Attachment Error

Incorrect: makes to rescue

Correct: rush to rescue

iii –

Prepositional Phrase Attachment Error

Incorrect: named to Midland

Correct: guy to Midland

iv –

Modifier Attachment Error

Incorrect: elements to most

Correct: crucial to most

Code Result

Average Train Loss: 0.1518338288490971

Evaluating on dev set

125250it [00:00, 10466122.96it/s]

- dev UAS: 70.66

Average Train Loss: 0.06645396859808402

Final evaluation on test set

2919736it [00:00, 38020411.19it/s]

- test UAS: 89.08