3️⃣

Week 3

태그

cs224n-2023-lecture06-fancy-rnn.pdf

cs224n-2023-lecture05-rnnlm.pdf

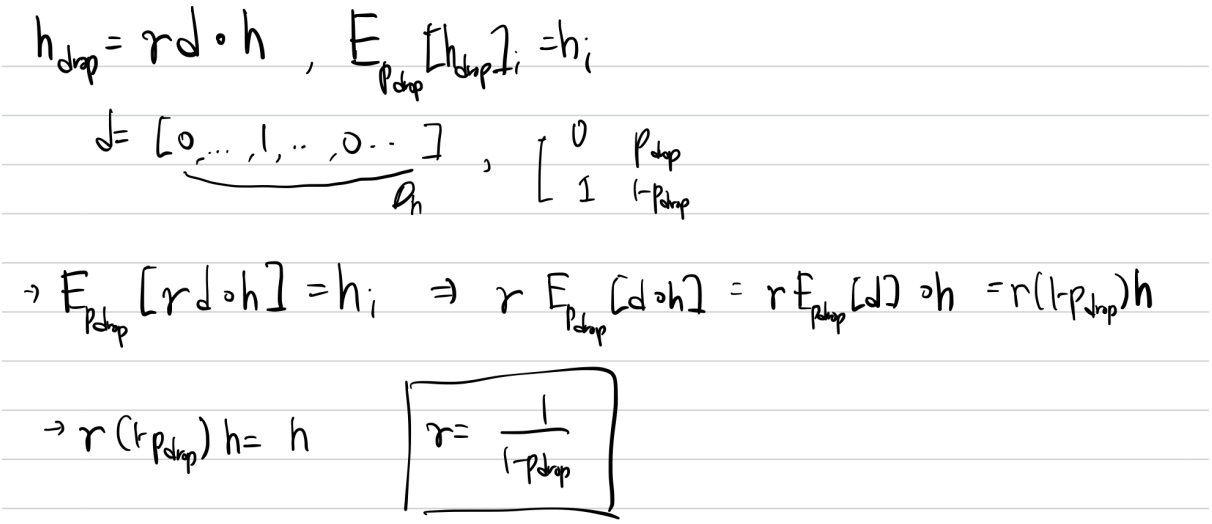
Homework 3

1

a-i: The usage of **m** provides low variance in updating θ by preserving the previous momentum value for a portion of β_1. This usage of momentum helps quick convergence by reducing the occurence of ociliations, and also reduces the influence of learning rate α.

a-ii: Because Adam divides the update by **v,** parameters with small **v's** will have larger updates. This division by v is an important point on Adam because it creates an adaptive learning rate. This adaptive learning rate helps the model parameters with different gradients converge with optimal speed.

b-i:



b-ii: Dropout technique is very helpful to prevent overfitting during training. It helps the model to gain additional generalization power. But it should not be applied during evaluation because all the information in the evaluation process should not be lost. Random dropout will also lead to inconsistent results, which is not desirable in evaluation process.

2

(a)

Stack	Buffer	New Dependency	Transition
[ROOT]	[I, attended, lectures, in, the, NLP, class]		Initial Configuration
[ROOT, I]	[attended, lectures, in, the, NLP, class]		SHIFT
[ROOT, I, attended]	[lectures, in, the, NLP, class]		SHIFT
[ROOT, attended]	[lectures, in, the, NLP, class]	attended → I	LEFT-ARC
[ROOT, attended, lectures]	[in, the, NLP, class]		SHIFT

[ROOT, attended]	[in, the, NLP, class]	attended → lectures	RIGHT-ARC
[ROOT, attended, in]	[the, NLP, class]		SHIFT
[ROOT, attended, in, the]	[NLP, class]		SHIFT
[ROOT, attended, in, the, NLP]	[class]		SHIFT
[ROOT, attended, in, the, NLP, class]	[]		SHIFT
[ROOT, attended, in, the, class]	[]	class → NLP	LEFT-ARC
[ROOT, attended, in, class]	[]	class → the	LEFT-ARC
[ROOT, attended, class]	[]	class → in	LEFT-ARC
[ROOT, attended]	[]	attended → class	RIGHT-ARC
[ROOT]	[]	ROOT → attended	RIGHT-ARC

(b)

The steps would be 2n because all n words goes through the stages of SHIFT and ARC (left or right).

Coding Result