Solution to cs224 assignment3(written)

Mhttx

February 27, 2019

1 Machine Learning and Neural Networks

1. a(i):

   Momentum m is the exponential smothing of gradients, hence reduce the variance;

   Momentum term increases for the dimensions whose gradients point in the same direction and reduces updates for dimensions whose gradients change directions, result in faster convergence and reduced oscillatin; avoiding getting trapped in their numerous suboptimal local minima(where gradient is 0)

2. a(ii):

   Parameters updated infrequently get large updates;

   Adam computes adaptive learning rates for each parameter, which enable us to update parameters according their crrosponding featrue's frenquencies, ie. perform a larger update for rarely occurring features and relative smaller update for frequently occurring features.

3. b(i):

$$\mathbb{E}_{p_{drop}}[\mathbf{h}_{drop}]_i = \gamma(1 - p_{drop})h_i = h_i \Rightarrow \gamma = \frac{1}{1 - p_{drop}}$$

4. b(ii)

   Dropout can be viewed as model ensembling, we train a single model with dropout and ensemble them during evaluation by not applying dropout.

2 Neural Transition-Based Dependency

1. (a):

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 → I	LEFT-ARC
[ROOT, parsed, this]	[sentence, correctly]		SHIFT
[ROOT, parsed, this, sentence]	[correctly]		SHIFT
[ROOT, parsed, sentence]	[correctly]	sentence → this	LEFT-ARC
[ROOT, parsed]	[correctly]	parsed → sentence	RIGHT-ARC
[ROOT, parsed, correctly]	[]		SHIFT
[ROOT, parsed]	[]	correctly → parsed	LEFT-ARC
[ROOT]	[]	parsed → ROOT	LEFT-ARC

2. (b):
   $2n$ steps, n steps for SHIFT words into stack and n *-ARC (RIGHT-ACR or LEFT-ARC) steps for moving words out from stack.