

CS 224n Assignment #3

Yoshihiro Kumazawa

June 1, 2020

1. Machine Learning & Neural Networks

- (a)
 - i. When β_1 is large, \mathbf{m} relies more on the history of the past gradients rather than the new one. For example, if $\beta_1 = 0.9$, the contribution of the new gradient to weight update is only 10% of that without momentum.
 - ii.
- (b)
 - i.
 - ii.

2. Neural Transition-Based Dependency Parsing

- (a)
- (b)
- (c)
- (d)
- (e)
- (f)