(a)

i. $m_{t+1} \leftarrow \beta_1 m_t + (1-\beta_1) \nabla_{\theta_t} J_{minibatch}(\theta_t)$

$\theta_{t+1} \leftarrow \theta_t - \alpha m_{t+1}$

① "m_t" contains the weighted previous gradient, with a high weight on current gradient, low weight on historical gradient (exponential decay)

② previous gradient may help the optimization process go through the saddle point and local optimal solutions.

ii. $m_{t+1} \leftarrow \beta_1 m_t + (1-\beta_1) \nabla_{\theta_t} J_{minibatch}(\theta_t)$

$V_{t+1} \leftarrow \beta_2 V_t + (1-\beta_2) \nabla_{\theta_t} J_{minibatch}(\theta_t)^2$

$\theta_{t+1} \leftarrow \theta_t - \alpha m_{t+1} / \sqrt{V_{t+1}}$

① The parameters with smaller gradient would have a larger update.

② It could prevent the model from local minima.

(b)

$h_{drop} = \gamma\, d \odot h \quad$ w.r.t. $E_{p_{drop}}[h_{drop}]_i = h_i$

i. $E_{p_{drop}}[h_{drop}]_i = p_{drop} \cdot 0 + (1-p_{drop}) \cdot h_i \cdot \gamma = h_i$

$\Rightarrow \gamma = \frac{1}{1-p_{drop}}$

ii. Served as regularization for model, because the decay of model complexity. During evaluation, we want the full information and averaged prediction of subnet.

2. (a)

Stack	Buffer	New dependency	Transition
[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) $n \cdot \text{shift} + n \cdot \text{Arc} = 2n$ times

(f) (i)

Error Type: Verb Phase Attachment error

Incorrect dependency: citing → acquisition

Correct dependency: citing → blocked

(ii) Error Type: Modifier Attachment error

Incorrect dependency: left → early

Correct dependency: afternoon → early

(iii) Error Type: Prepositional Attachment Error

Incorrect dependency: declined → decision

Correct dependency: reason → decision

(iv) Error Type: Coordination Attachment Error

Incorrect dependency: affects → one

Correct dependency: plants → one

(g) embeding → train_x

make more semantic sense.