

a3

一、

a.i

使用m（动量）可以帮助平滑梯度更新，因为它累积了之前梯度的滚动平均值，避免了由于单次梯度波动导致的过大更新。这样可以减少梯度的方差，使得参数更新更加稳定，有助于模型更快收敛并避免陷入局部极小值，从而提升整体学习效果。

a.ii

在 Adam 算法中，由于更新量被 \sqrt{v} 除，那些具有较小梯度波动（或梯度较小）的参数会得到相对更大的更新，而梯度较大的参数更新幅度会相对较小。这种自适应的学习率可以防止较大的梯度引发剧烈的参数波动，同时加速较小梯度参数的更新，帮助模型更快地收敛到最优解，并提升训练的稳定性和效率。

b.i

解答：我们需要通过 γ 来缩放输出，以确保测试时的输出尺度与训练时的期望输出一致。如果不乘以 γ ，我们可以看到在训练过程中任何输出项的期望值为：

$$\mathbb{E}_{p_{drop}}[\mathbf{h}_{drop}]_i = p_{drop} \cdot 0 + (1 - p_{drop})h_i = (1 - p_{drop})h_i$$

因此，在测试时没有任何单元被丢弃，我们需要将输出向量乘以 $(1 - p_{drop})$ ，以匹配训练时的期望值。为了保持测试不变，我们可以在训练期间应用反向 dropout，通过乘以预测期间应乘数的倒数来调整输出。我们可以推导出：

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

b.ii

在训练过程中，应用 dropout 有助于防止过拟合。它通过随机丢弃一些神经元，使模型不依赖于特定的神经元，从而促使模型学习更具鲁棒性的特征表示。这种随机性可以提高

模型的泛化能力，因为它有效地相当于在训练多个不同的子网络。

在评估过程中，不应使用 dropout，因为我们希望模型的表现具有确定性，所有神经元都应被激活以充分利用模型的学习能力。此外，在评估时禁用 dropout 可以确保输出的一致性，使得模型能够使用在训练过程中学习到的完整表示，避免不必要的随机性带来的性能波动。

二、

a.

"I attended lectures in the NLP class" 所需的转换序列：

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.

解答：包含 n 个单词的句子将在 $2n$ 步内解析完成，因为依存解析器的时间复杂度是线性的。换句话说，它处理的是成对的单词，不支持嵌套依存关系。因此，完成所有 SHIFT 操作需要 n 步，完成所有 LEFT-ARC 和 RIGHT-ARC 操作则需要另外 n 步。

e.

开发集上最佳UAS为88.15

测试集上UAS为88.90

- **正确的依存关系**：afternoon → early

(iii)

- **错误类型**：介词短语附着错误 (Prepositional Phrase Attachment Error)
- **错误的依存关系**：declined → decision
- **正确的依存关系**：reason → decision

(iv)

- **错误类型**：并列关系附着错误 (Coordination Attachment Error)
- **错误的依存关系**：affects → one
- **正确的依存关系**：plants → one

g.

1. **提供句法信息**：词性标注能够捕捉单词的句法角色，例如区分名词、动词或介词等，从而帮助解析器更好地理解单词在句子中的功能。
2. **减少歧义**：某些单词可能具有多种含义（如名词或动词），而词性标注可以消除这种歧义，提供更准确的上下文信息。
3. **增强泛化能力**：当解析器遇到未见过的单词时，词性标注可以作为一种高层次的抽象信息，帮助解析器推断依存关系，尤其是在小数据集或低频词的情况下。
4. **提高依存关系准确性**：通过结合词性标注，解析器可以更容易识别句法模式，例如主谓宾结构，从而提升解析性能。