IFT6289-A-H26 Assignment Report

Student Information
Student Name: _____
Student ID: _____
Assignment: _____
Date: _____

Part 1: Derive the gradients for training

1.1 Naïve softmax: $\partial \mathbf{L}_{\mathbf{naive}}/\partial \boldsymbol{v}_i$

1.2 Naïve softmax: $\partial \mathbf{L}_{\mathbf{naive}}/\partial \boldsymbol{u}_w$ **(two cases** $w = o$ **and** $w \neq o$**)**

1.3 Negative sampling: $\partial \mathbf{L}_{\mathbf{neg\_sampling}}/\partial \boldsymbol{v}_i$

1.4 Negative sampling: $\partial \mathbf{L}_{\mathbf{neg\_sampling}}/\partial \boldsymbol{u}_o$

1.5 Negative sampling: $\partial \mathbf{L}_{\mathbf{neg\_sampling}}/\partial \boldsymbol{u}_k$

Part 3: Analyzing your word vectors

3.1 Training results (include plots + analysis)
3.1(a) Expectations: What characteristics do you expect (e.g., clusters, analogies/arithmetic)?

3.1(b) Observations: Are the results in line with your expectations? Refer to the plots.

3.2 Penn Treebank preprocessing analysis

3.2(a) ⟨UNK⟩: Pros/cons of keeping top 10K words; ideas to reduce ⟨UNK⟩tokens with low overhead.

3.2(b) Numbers → N: Pros/cons; ideas to handle numbers reasonably.

References

If you cite papers, add a `references.bib` file to the project and use `\citep{...}` / `\citet{...}`.