# CSE 143 Assignment 2

Liam Xu, Jaisuraj Kaleeswaran, Vishwa Vijayasankar, Anthony Reyna

April 2024

## 1 Programming: n-gram language modeling

Training set

| Model | Perplexity |
|---|---|
| Unigram | 862.60 |
| Bigram | 55.11 |
| Trigram | Not implemented |

Dev set

| Model | Perplexity |
|---|---|
| Unigram | 788.29 |
| Bigram | 19.33 |
| Trigram | Not implemented |

## 2 Programming: additive smoothing

Training set, $\alpha = 1$

| Model | Perplexity |
|---|---|
| Unigram | 16.27 |
| Bigram | Not implemented |
| Trigram | Not implemented |

Training set, $\alpha = 5$

| Model | Perplexity |
|---|---|
| Unigram | 15.52 |
| Bigram | Not implemented |
| Trigram | Not implemented |

Training set, $\alpha = 10$

| Model | Perplexity |
|---|---|
| Unigram | 14.84 |
| Bigram | Not implemented |
| Trigram | Not implemented |

Dev set, $\alpha = 1$

| Model | Perplexity |
|---|---|
| Unigram | 14.89 |
| Bigram | Not implemented |
| Trigram | Not implemented |

Dev set, $\alpha = 5$

| Model | Perplexity |
|---|---|
| Unigram | 14.25 |
| Bigram | Not implemented |
| Trigram | Not implemented |

Dev set, $\alpha = 10$

| Model | Perplexity |
|---|---|
| Unigram | 13.67 |
| Bigram | Not implemented |
| Trigram | Not implemented |