CS 533: Natural Language Processing **(Due: 01/31/20)**

Assignment 1

Name: your_name, *NetID:* your_id *Students discussed with:* names_of_students

Problem 1: Preliminaries

((1+1+1+1) + (1+1+1+1) + (1+1+1) = 11 points)

1. **(Probability)**
   (a) Answer
   (b) Answer
   (c) Answer
   (d) Answer
2. **(Linear Algebra)**
   (a) Answer
3. **(Optimization)**
   (a) Answer
   (b) Answer
   (c) Answer

Problem 2: n-Gram Models

(4 + (2+1+1+1) = 9 points)

1. **(Relative Frequency Lemma)**
   (a) Answer
2. **(Maximum Likelihood Estimation (MLE) of the Trigram Language Model)**
   (a) Answer
   (b) Answer
   (c) Answer
   (d) Answer

Problem 3: Programming

(2 + 1 + 1 + 1 + 2 + 1 + 3 + 1 = 12 points)

(Code must be submitted as well, with unambiguous commands for replicating reported results.)

1. Answer
2. Answer
3. Answer
4. Answer
5. Answer
6. Answer
7. Answer
8. Answer

Advice on writing in LaTeX (do not include this part in your submission) When you display equations, use `align` for numbered equations and `align*` for unnumbered equations. For instance, here is an unnumbered equation on display:

$$f(x) = Wx$$

Here are numbered equations on display:

$$\begin{aligned} f(x) &= Wx \\ g(y) &= \tanh(y) \\ h(x) &= g(f(x)) \end{aligned} \tag{1}$$

Refer to specific equations by using `eqref`: "equation (1)". Use `$$` to write inline equations: $f(x) = Wx$. Enforce consistency by using predefined commands. Take a look at commands defined in `header.tex`. For instance, you can write non-default operations like `argmax`:

$$x^* = \arg\max_{x \in \mathbb{R}} f(x)$$

You can automatically adjust heights of brackets/parentheses by using `expected`:

$$D_{\mathrm{KL}}(p_X || q_X) = \mathop{\mathbf{E}}_{x \sim p_X} \left[\log \left(\frac{p_X(x)}{q_X(x)} \frac{r_X(x)}{r_X(x)} \right) \right]$$

Use `ldots` to write ..., not three periods. Do not confuse $\sum$ (summation) with Σ (Greek letter), or $\prod$ (product) with Π (Greek letter).