CSE556: NLPQuiz - 2

Duration: 40 mins + **5 mins for uploading.**

Total marks: 30

Note:

- It's a pen-and-paper-based quiz. **On every sheet, please write your name in BOLD LETTERS.** We will not evaluate if you fail to do so.
- There will not be any extension and also no exemption will be given to anyone.

1. "Smoothing always incurs some information loss to the original counts". Justify or refute. **[3]**

2. Let V = {a, b, w}, Corpus: [{"a b b a a w a b w b"}, {"w a a a b a"}]. Utilizing the trigram LM, find the next predicted word in the sequence "w b a ____". Use **Interpolation** as the smoothing technique, in case you're getting zero co-occurrence counts. Take appropriate alpha values. **Show computation**. **[15]**

3. Write shorthand orthographic rules (similar to the equation below) for Y replacement -- "*-y* changes to *-ie* before *-s*, *-i* before *-ed*". **[4]**

$$FST_i: \quad \varepsilon \rightarrow e \, / \left\{ \begin{array}{c} x \\ s \\ z \end{array} \right\} \char`\^ \underline{\quad} s\#$$

[No partial marking.]

4. Write the **lexical** and **intermediate** forms for the surface forms **Boys**, **Cities**, and **Fish**. Use conventional notations. **[8]**