Quiz 1 Answer Schema

Q1. What are the factors affecting the Discrimination and Reliability of a Language Model? How?

<u>Answer:</u>

1. The primary cause is the length of the horizon.

2. How-> Larger n-grams are rarer resulting in less reliability but carrying a lot of information resulting in better discrimination.

Q2. How will you compare two language models trained on the same data? (How will you say which one is better?)

<u>Answer:</u>

1. Perplexity over the same data. Explain in detail

2. Bonus for other unique ideas.

Q3. What are the ways in which Language Models can avoid sparsity?

<u>Answer:</u>

1. Avoiding sparsity is impossible :) Bonus if someone says this.

2. Smoothing and back-off.

Q4. How can multiple Language models be combined to get better Language models? Give some examples of the models you know.

<u>Answer:</u>

1. Interpolation and back-off.

2. Model examples and explanations.

_____