Task 2 Report

1. Top 5 bigrams

BEFORE SMOOTHING
   a) ('href', 'http'), score = 1
   b) ('tychelle', 'to'), score = 1
   c) ('hang', 'out'), score = 1
   d) ('nonexistent', 'social'), score = 1
   e) ('alex', 'and'), score = 1

AFTER LAPLACE SMOOTHING
   a) ('i', 'feel'), score = 0.110436
   b) ('feel', 'like'), score = 0.035097
   c) ('i', 'am'), score = 0.031894
   d) ('that', 'i'), score = 0.026506
   e) ('and', 'i'), score = 0.023103

AFTER KNESER-NEY SMOOTHING
   a) ('href', 'http'), score = 0.97200218
   b) ('don', 't'), score = 0.9712049203427449
   c) ('didn', 't'), score = 0.9611413972283877
   d) ('sort', 'of'), score = 0.9594087640897253
   e) ('supposed', 'to'), score = 0.9238243578261

2. Reason for Beta Emotion component

The main intuition behind the addition of the beta emotion component is derived from the way cross-entropy loss works. We created an additional 2D matrix/dictionary (nested) for bigram words to store vectors of size 6, i.e., R^6 (like the way we create to store the bigram count).

For each bigram in the sentence, we added the six-sized emotion vector (this we obtained from emotion_score function). This was repeated for every sentence in the corpus. This matrix of R^6 vectors allowed us to select the desired emotion values to be tweaked and, hence the particular emotion.

When asked to generate an emotion sentence, firstly, emotion was one-hot encoded to R^6 array. Afterward, for each token (that can be a possible token after context), the beta score was calculated as the product of the one-hot vector and the emotion vector of context-token bigram. Finally, the normalized beta score

was tweaked into the probability formula (with both addition and multiplication terms for better results). And the token was probabilistically sampled from these probabilities.

The advantage of using the matrix of the R^6 vector was that it allowed us to select the emotion component that we wanted and, hence to generate a sentence specific for that emotion only.

3. Add two generated samples for each emotion

For Anger:
1. i feel that i feel offended by their souls
2. i feel a little frustrated and the work place wore glasses for no reason

For Fear:
1. i feel threatened by and collected. i feel weird
2. i feel unsure of being negative right now

For Joy:
1. i feel the clay and enjoy feeling fine tonight
2. i feel inspired to be a habanero if i feel the beginning to be brave

For Love:
1. i feel when i feel so supportive of him i feel like sweet of me and i feel tender and that i feel accepted there for each other books when
2. i feel the sweet. i feel like hes the one of your fragrance

For Sadness:
1. i feel very dull and work is stupid and tv or the world
2. i feel very regretful when i feel abused

For Surprise:
1. i feel surprised at the ground floor feeling this is weird pressure of my own lounge room unless i feel the internet
2. i feel funny. i feel amazed imagining how much more amazed because i feel funny

4. Accuracy and macro F1 scores obtained from extrinsic evaluation

Grid search for differently parameterized SVC classifiers gave us the best parameter of linear kernel and unit regularization parameter.
Below are the scores on scale of 1
Accuracy score on Testing set corpus: 0.9812
Macro F1 score on Testing set corpus: 0.98121
Accuracy score on Testing set generated: 0.98
Macro F1 score on Testing set generated: 0.9799278

5. For each emotion, pick 1 of the generated sample and reason why it is generated according to its corresponding emotion.

For Anger:
Sentence: i feel that i feel offended by their souls
Reason: For anger emotion, words related to anger (offended, etc) were selected by the product of one-hot emotion vector and emotion vector from matrix. Zero terms in one-hot vector nullified the possibility of other words not relevant to anger emotion, and probabilistically sampled got picked.

For Fear:
Sentence: i feel threatened by and collected. i feel weird
Reason: Same reason as above, but this time, the one-hot emotion vector will only select/increase probability for tokens related to fear and will nullify other tokens.

For Joy:
Sentence: i feel inspired to be a habanero if i feel the beginning to be brave
Reason: Same reason as above, but this time, the one-hot emotion vector will only select/increase probability for tokens related to joy and will nullify other tokens.

For Love:
Sentence: i feel the sweet. i feel like hes the one of your fragrance
Reason: Same reason as above, but this time, the one-hot emotion vector will only select/increase probability for tokens related to love and will nullify other

For Sadness:
Sentence: i feel very regretful when i feel abused
Reason: Same reason as above, but this time, the one-hot emotion vector will only select/increase the probability for tokens related to sadness and will nullify other

For Surprise:
Sentence: i feel funny. i feel amazed imagining how much more amazed because i feel funny
Reason: Same reason as above, but this time, the one-hot emotion vector will only select/increase probability for tokens related to surprise and will nullify other

6. Credit Statement:

Individual contributions are as given below.

Khushdev Pandit
Attempted Task 1.
Also attempted Q1, Q2 and Q3 of Task 2 with the intuition of beta emotion (the one described above) different from my teammates.

Apurv Dube
Attempted Task 2, parts 1, 2 and 3. The code and results are attached.

Pankaj
Attempted Task 2, part 3 for improvement and part 4. The code is attached.

Sara Moin
Attempted Task 1.