CS 411 - Homework 1 - Ege Demirci

Question - 1

The code uses a very basic shift-cipher approach. It tries to decrypt the given ciphertext "NLPDLC" for all possible shift values from 1 to 25 (since a shift of 0 would result in the original ciphertext). It calls decrypt_shift function for each key and prints the decrypted word along with the corresponding key.

The meaningful word I found in this question is "Caesar", and this word was found when the key is 11.

Question - 2

First I took some functions from helper_py file. I then computed the frequency of each character in the ciphertext using `Counter` class, so as to identify the character that possibly maps to 'T' in the plaintext. As it said in the lecture, frequency analysis often gives clues about possible character mappings in classical ciphers.

Then I implemented the `decrypt_affine` function which decrypts a given ciphertext using the Affine Cipher formula, given values for `a` and `b`.

The decryption formula is: $P = (a_{\text{inv}} \cdot (C - b)) \mod 26$

Where P is the plaintext character,

a_{inv} is the modular inverse of a,

C is the ciphertext character, and

b is the shift.

Knowing that the Affine Cipher involves a multiplication and a shift, I first considered possible values for `a` that have modular inverses in modulo 26. These are $[1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23, 25]$.

For each `a` value, I derived a `b` value based on the hint that the most frequent letter maps to 'T'. So, I used the formula:

$$b = (\text{lowercase}('a') - \text{lowercase}('t') \cdot a) \mod 26$$

for the one of the most common characters 'a' from the ciphertext. I also derived another `b` value for the other most common character 'r' from the ciphertext

using the formula:

$$b = (\text{lowercase}('r') - \text{lowercase}('t') \cdot a) \mod 26$$

. I then tried decrypting the ciphertext using each combination of `a` and `b`.

Among the decrypted texts, the one which was meaningful: "I did not fail the test. I just found three hundred eleven ways to do it wrong." This was decrypted using $a = 11$ and $b = 25$.

Question - 3

When employing single characters, the modulus is naturally 30, given the 30 distinct symbols in our set. However, when we are using bigraphs, we're essentially combining each of the 30 characters with every other character, including itself. This results in $30 \times 30 = 900$ possible combinations. Therefore, the modulus for the Affine Cipher with bigraphs is 900.

As I explained in the previous question, for the affine cipher, the key comprises two values, a and b.

The value for a must be chosen such that it's relatively prime to 900, ensuring it has a modular inverse. There are 240 such potential values for a, and I showed this in the notebook by counting the elements.

For b, any value between 0 and 899 is valid, yielding 900 potential choices. Thus, the size of the key space is $240\,(a\text{-values}) \times 900\,(b\text{-values}) = 216,000$ possible keys.

Question - 4

No, the Affine Cipher defined for two-letter groups is not inherently secure against letter frequency analysis. Firstly, it's true that using bigrams can provide a larger set of possible encodings, but the underlying frequency patterns of bigrams in the English language remain consistent. The frequency analysis assumes that there are individual letters in English that appear more frequently than others, but there are certain bigrams that tend to occur more often as well. For instance, bigrams such as "th", "he", "in", "en", "re", and "er" are among the most commonly used in the English language.

If an attacker is aware of these patterns, they can use frequency analysis to identify the most common bigrams in the ciphertext and potentially map them to their plaintext counterparts.

Therefore, even though this modified Affine Cipher encrypts two letters at a time, it remains susceptible to attacks based on bigram frequency analysis.

Question - 5

First of all, I have defined a function `BigramEncrypt` that, given a bigram, computes its corresponding numerical value using the provided alphabet just like

in the question. Then, I populated the `bigrams` and `inv_bigrams` dictionaries for each combination. These dictionaries are important for converting between the bigram representation (like 'TH') and its numerical encoding (like 577). The `decrypt` function, again just like in the previous questions, decrypts a ciphertext given encryption keys a and b, and modulus m. This function works by converting each bigram from the ciphertext into its numerical representation, decrypting it using the affine decryption formula, and then mapping the result back to its bigram form. But for this purpose, we needed alpha and beta values. With the help of the given hint, we know that the plaintext ends with a ".X", since if the number of letters in the plaintext is not a multiple of two, we pad it with the letter "X" at the end.

I extracted the last bigram from the ciphertext ('SW') and used it to derive a potential value for b based on a guessed value for a and the encoded value for the bigram '.X'.

I used the following formulation:

$$\text{bigrams}['SW'] \equiv a \times \text{bigrams}['.X'] + b \mod 900$$

From this, we can deduce b as:

$$b \equiv \text{bigrams}['SW'] - a \times \text{bigrams}['.X'] \mod 900$$

With the modulus being 900 and knowing the properties of the Affine Cipher, I iterated through all possible values of a which are coprime to 900. For each of these values, I computed a potential b value using the equation derived from the hint.

For each (a, b) key pair, I attempted decryption and checked the resulting plaintext.

The decryption process resulted in the meaningful plaintext: "SING, GODDESS, OF THE ANGER OF ACHILLES, SON OF PELEUS." using the key pair $a = 91$ and $b = 389$.

Question - 6

For this question, we can prove in the following way:

Let's consider a specific letter α from the Turkish alphabet just assumed in the question. Given that we are shifting by a random and uniform value, the probability of α becoming any specific letter β in the ciphertext is $\frac{1}{29}$, since there are 29 letters in the alphabet.

The probability of getting the ciphertext letter β given the plaintext letter α. Since the shift is uniform and random:

$$P(\beta|\alpha) = \frac{1}{29}$$

If we compute the overall probability of getting the ciphertext letter β by summing over all possible plaintext letters:

$$p_\beta = \sum_\alpha P(\beta|\alpha) \times p_\alpha$$

Given our second point, $P(\beta|\alpha)$ is $\frac{1}{29}$ for all α.
Therefore, the equation becomes:

$$p_\beta = \sum_\alpha \frac{1}{29} \times p_\alpha = \frac{1}{29}$$

So, no matter how p_α is distributed, p_β is always $\frac{1}{29}$ for every letter β in the Turkish alphabet.
Q.E.D.

Question - 7

I started by researching the Vigènere cipher, and I couldn't determine the key length as we saw in the class. When I researched on the internet, I found the Kasiski technique and proceeded from there. What I did was to identify sequences repeating at a certain length. By identifying such sequences and noting the distances between their occurrences, I tried to reach the length of the encryption key. Before proceeding, I removed non-letter characters from the encrypted text by focusing only on encrypted alphabetical characters. This was necessary because the Vigènere cipher here only works on letters, and including non-letter characters could corrupt our analysis. After obtaining a list of the distances between repeated sequences, I tried to determine the most common factors of these distances using the `find_common_factors` function. The most frequently observed factors can give clues about the possible lengths of the key. This step is primarily based on the idea behind the Kasiski examination; this idea suggests that if a particular key length is correct, the repeated sequences in the encrypted text may correspond to the same plaintext sequences encrypted by the same part of the key. After calculating the common factors, I ranked them according to their frequency. The most common factors were 5 and 2. Since 2 is too short for a key length, I continued my analysis with 5. Now that I had an idea about the possible key length, I proceeded to the frequency analysis of the cipher text to determine the individual shifts used in the Vigènere cipher.

I created a dictionary (`letter_count`) to count the occurrence of each letter in a portion of the encrypted text. By comparing the most frequently observed letter in a portion of the encrypted text with the most commonly used letter in the English language (usually 'E'), I was able to estimate the shift used for that section. The `Find_shift` function was exactly for this. From here, I reached the key MGVDQ. When I used the MGVDQ key, I noticed that some words couldn't be decrypted precisely, so I started trying brute-force for the last word of the key and printed all the elements from A to Z, and saw that the MGVDB key yielded meaningful output.

And I reached the decrypted message:

THE CENTRIPETAL FORCE ON OUR PLANET IS STILL FEARFULLY STRONG, ALYOSHA. I HAVE A LONGING FOR LIFE, AND I GO ON LIVING IN SPITE OF LOGIC. THOUGH I MAY NOT BELIEVE IN THE ORDER OF THE UNIVERSE, YET I LOVE THE STICKY LITTLE LEAVES AS

THEY OPEN IN SPRING. I LOVE THE BLUE SKY, I LOVE SOME PEOPLE, WHOM ONE LOVES YOU KNOW SOMETIMES WITHOUT KNOWING WHY. I LOVE SOME GREAT DEEDS DONE BY MEN, THOUGH I'VE LONG CEASED PERHAPS TO HAVE FAITH IN THEM, YET FROM OLD HABIT ONE'S HEART PRIZES THEM. HERE THEY HAVE BROUGHT THE SOUP FOR YOU, EAT IT, IT WILL DO YOU GOOD. IT'S FIRST-RATE SOUP, THEY KNOW HOW TO MAKE IT HERE. I WANT TO TRAVEL IN EUROPE, ALYOSHA, I SHALL SET OFF FROM HERE. AND YET I KNOW THAT I AM ONLY GOING TO A GRAVEYARD, BUT IT'S A MOST PRECIOUS GRAVEYARD, THAT'S WHAT IT IS. PRECIOUS ARE THE DEAD THAT LIE THERE, EVERY STONE OVER THEM SPEAKS OF SUCH BURNING LIFE IN THE PAST, OF SUCH PASSIONATE FAITH IN THEIR WORK, THEIR TRUTH, THEIR STRUGGLE AND THEIR SCIENCE, THAT I KNOW I SHALL FALL ON THE GROUND AND KISS THOSE STONES AND WEEP OVER THEM; THOUGH I'M CONVINCED IN MY HEART THAT IT'S LONG BEEN NOTHING BUT A GRAVEYARD. AND I SHALL NOT WEEP FROM DESPAIR, BUT SIMPLY BECAUSE I SHALL BE HAPPY IN MY TEARS, I SHALL STEEP MY SOUL IN EMOTION. I LOVE THE STICKY LEAVES IN SPRING, THE BLUE SKY - THAT'S ALL IT IS. IT'S NOT A MATTER OF INTELLECT OR LOGIC, IT'S LOVING WITH ONE'S INSIDE, WITH ONE'S STOMACH.