Khandelwal-HW1

September 25, 2024

1 Problem 1

[7]:
```
!pip install -q transformers evaluate
```

[2]:
```
from transformers import GPT2LMHeadModel, GPT2TokenizerFast
import torch
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import warnings
from tqdm import tqdm

warnings.filterwarnings('ignore')
import os
os.environ['CUDA_LAUNCH_BLOCKING'] = "1"
os.environ['TORCH_USE_CUDA_DSA'] = '1'
```

[5]:
```
device = "cuda" if torch.cuda.is_available() else "cpu"
```

[4]:
```
test = pd.read_csv('/kaggle/input/data-hw1-twai/Problem1_dataset.csv')
```

1.1 a)

[5]:
```
print('First 10 test samples:')
for i, example in enumerate(test.text.values[:10]):
  print(f'{i+1}. {example}')
```

```
First 10 test samples:
1.  = Robert Boulter =

2.  Robert Boulter is an English film , television and theatre actor . He had a
guest @-@ starring role on the television series The Bill in 2000 . This was
followed by a starring role in the play Herons written by Simon Stephens , which
was performed in 2001 at the Royal Court Theatre . He had a guest role in the
television series Judge John Deed in 2002 . In 2004 Boulter landed a role as "
Craig " in the episode " Teddy 's Story " of the television series The Long Firm
; he starred alongside actors Mark Strong and Derek Jacobi . He was cast in the
2005 theatre productions of the Philip Ridley play Mercury Fur , which was
```

performed at the Drum Theatre in Plymouth and the Menier Chocolate Factory in London . He was directed by John Tiffany and starred alongside Ben Whishaw , Shane Zaza , Harry Kent , Fraser Ayres , Sophie Stanton and Dominic Hall .

3. In 2006 , Boulter starred alongside Whishaw in the play Citizenship written by Mark Ravenhill . He appeared on a 2006 episode of the television series , Doctors , followed by a role in the 2007 theatre production of How to Curse directed by Josie Rourke . How to Curse was performed at Bush Theatre in the London Borough of Hammersmith and Fulham . Boulter starred in two films in 2008 , Daylight Robbery by filmmaker Paris Leonti , and Donkey Punch directed by Olly Blackburn . In May 2008 , Boulter made a guest appearance on a two @-@ part episode arc of the television series Waking the Dead , followed by an appearance on the television series Survivors in November 2008 . He had a recurring role in ten episodes of the television series Casualty in 2010 , as " Kieron Fletcher " . Boulter starred in the 2011 film Mercenaries directed by Paris Leonti .

4. = = Career = =

5. = = = 2000 - 2005 = = =

6. In 2000 Boulter had a guest @-@ starring role on the television series The Bill ; he portrayed " Scott Parry " in the episode , " In Safe Hands " . Boulter starred as " Scott " in the play Herons written by Simon Stephens , which was performed in 2001 at the Royal Court Theatre . A review of Boulter 's performance in The Independent on Sunday described him as " horribly menacing " in the role , and he received critical reviews in The Herald , and Evening Standard . He appeared in the television series Judge John Deed in 2002 as " Addem Armitage " in the episode " Political Expediency " , and had a role as a different character " Toby Steele " on The Bill .

7. He had a recurring role in 2003 on two episodes of The Bill , as character " Connor Price " . In 2004 Boulter landed a role as " Craig " in the episode " Teddy 's Story " of the television series The Long Firm ; he starred alongside actors Mark Strong and Derek Jacobi . Boulter starred as " Darren " , in the 2005 theatre productions of the Philip Ridley play Mercury Fur . It was performed at the Drum Theatre in Plymouth , and the Menier Chocolate Factory in London . He was directed by John Tiffany and starred alongside Ben Whishaw , Shane Zaza , Harry Kent , Fraser Ayres , Sophie Stanton and Dominic Hall . Boulter received a favorable review in The Daily Telegraph : " The acting is shatteringly intense , with wired performances from Ben Whishaw (now unrecognisable from his performance as Trevor Nunn 's Hamlet) , Robert Boulter , Shane Zaza and Fraser Ayres . " The Guardian noted , " Ben Whishaw and Robert Boulter offer tenderness amid the savagery . "

8. = = = 2006 - present = = =

9. In 2006 Boulter starred in the play Citizenship written by Mark Ravenhill . The play was part of a series which featured different playwrights , titled Burn

```
/ Chatroom / Citizenship . In a 2006 interview , fellow actor Ben Whishaw
identified Boulter as one of his favorite co @-@ stars : " I loved working with
a guy called Robert Boulter , who was in the triple bill of Burn , Chatroom and
Citizenship at the National . He played my brother in Mercury Fur . " He
portrayed " Jason Tyler " on the 2006 episode of the television series , Doctors
, titled " Something I Ate " . Boulter starred as " William " in the 2007
production of How to Curse directed by Josie Rourke . How to Curse was performed
at Bush Theatre in the London Borough of Hammersmith and Fulham . In a review of
the production for The Daily Telegraph , theatre critic Charles Spencer noted ,
" Robert Boulter brings a touching vulnerability to the stage as William . "

10.  Boulter starred in two films in 2008 , Daylight Robbery by filmmaker Paris
Leonti , and Donkey Punch directed by Olly Blackburn . Boulter portrayed a
character named " Sean " in Donkey Punch , who tags along with character " Josh
" as the " quiet brother … who hits it off with Tammi " . Boulter guest
starred on a two @-@ part episode arc " Wounds " in May 2008 of the television
series Waking the Dead as character " Jimmy Dearden " . He appeared on the
television series Survivors as " Neil " in November 2008 . He had a recurring
role in ten episodes of the television series Casualty in 2010 , as " Kieron
Fletcher " . He portrayed an emergency physician applying for a medical
fellowship . He commented on the inherent difficulties in portraying a physician
on television : " Playing a doctor is a strange experience . Pretending you know
what you 're talking about when you don 't is very bizarre but there are
advisers on set who are fantastic at taking you through procedures and giving
you the confidence to stand there and look like you know what you 're doing . "
Boulter starred in the 2011 film Mercenaries directed by Paris Leonti .
```

[6]:

```python
# Print dataset statistics:
samples = test.text.values
total_samples = len(samples)
print(f'There are {total_samples} samples!\n')

total_chars = len("".join([e for e in samples]))

for i,sample in enumerate(samples):
  if i==5:
    break
  avg_num_chars = round(len(sample)/total_chars,5)
  print('Sample - ', sample, end='\r')
  print(f'There are on average {avg_num_chars} characters in the {i+1} sample')
  print()
```

```
There are 187 samples!

Sample -   = Robert Boulter =
There are on average 0.00023 characters in the 1 sample
```

```
Sample -   Robert Boulter is an English film , television and theatre actor . He
had a guest @-@ starring role on the television series The Bill in 2000 . This
was followed by a starring role in the play Herons written by Simon Stephens ,
which was performed in 2001 at the Royal Court Theatre . He had a guest role in
the television series Judge John Deed in 2002 . In 2004 Boulter landed a role as
" Craig " in the episode " Teddy 's Story " of the television series The Long
Firm ; he starred alongside actors Mark Strong and Derek Jacobi . He was cast in
the 2005 theatre productions of the Philip Ridley play Mercury Fur , which was
performed at the Drum Theatre in Plymouth and the Menier Chocolate Factory in
London . He was directed by John Tiffany and starred alongside Ben Whishaw ,
Shane Zaza , Harry Kent , Fraser Ayres , Sophie Stanton and Dominic Hall .
There are on average 0.00949 characters in the 2 sample

Sample -   In 2006 , Boulter starred alongside Whishaw in the play Citizenship
written by Mark Ravenhill . He appeared on a 2006 episode of the television
series , Doctors , followed by a role in the 2007 theatre production of How to
Curse directed by Josie Rourke . How to Curse was performed at Bush Theatre in
the London Borough of Hammersmith and Fulham . Boulter starred in two films in
2008 , Daylight Robbery by filmmaker Paris Leonti , and Donkey Punch directed by
Olly Blackburn . In May 2008 , Boulter made a guest appearance on a two @-@ part
episode arc of the television series Waking the Dead , followed by an appearance
on the television series Survivors in November 2008 . He had a recurring role in
ten episodes of the television series Casualty in 2010 , as " Kieron Fletcher "
. Boulter starred in the 2011 film Mercenaries directed by Paris Leonti .
There are on average 0.00952 characters in the 3 sample

Sample -   = = Career = =
There are on average 0.00019 characters in the 4 sample

Sample -   = = = 2000 - 2005 = = =
There are on average 0.00029 characters in the 5 sample
```

1.2 b)

[7]:

```python
def calculate_perplexity(text, model, tokenizer, device, window_length=1024, step_size=256):

    # Tokenize the text
    tokens = tokenizer.encode(text, return_tensors='pt').to(device)  # Ensure tokens are on GPU
    # Get the length of tokens
    num_tokens = tokens.size(1)

    # Initialize total loss and count
    total_loss = 0
    count = 0
    # Move the model to the specified device (GPU)
    model.to(device)

    # Iterate over the text in windows
    for start in range(0, num_tokens, step_size):
        end = min(start + window_length, num_tokens)
        input_ids = tokens[:, start:end].to(device)  # Move input_ids to GPU

        # Ensure input_ids is not empty
        if input_ids.size(1) == 0:
            continue

        # Get model outputs and calculate loss
        with torch.no_grad():
            outputs = model(input_ids, labels=input_ids)
            loss = outputs.loss.item()

        # Accumulate loss and count
        total_loss += loss * input_ids.size(1)
        count += input_ids.size(1)

    # Calculate average loss
    average_loss = total_loss / count

    # Calculate perplexity
    perplexity = np.exp(average_loss)

    return perplexity
```

[8]:
```
# Complete code for HuggingFace GPT2 model (small, medium, and large)␣
 ↪evaluation, according to the perplexity metric

# Load Models
gpt2_small_model = GPT2LMHeadModel.from_pretrained('gpt2')
gpt2_small_tokenizer = GPT2TokenizerFast.from_pretrained('gpt2')

gpt2_medium_model = GPT2LMHeadModel.from_pretrained('gpt2-medium')
gpt2_medium_tokenizer = GPT2TokenizerFast.from_pretrained('gpt2-medium')

gpt2_large_model = GPT2LMHeadModel.from_pretrained('gpt2-large')
gpt2_large_tokenizer = GPT2TokenizerFast.from_pretrained('gpt2-large')
```

```
config.json:   0%|          | 0.00/665 [00:00<?, ?B/s]

model.safetensors:   0%|          | 0.00/548M [00:00<?, ?B/s]

generation_config.json:   0%|          | 0.00/124 [00:00<?, ?B/s]
```

```
tokenizer_config.json:   0%|          | 0.00/26.0 [00:00<?, ?B/s]

vocab.json:   0%|          | 0.00/1.04M [00:00<?, ?B/s]

merges.txt:   0%|          | 0.00/456k [00:00<?, ?B/s]

tokenizer.json:   0%|          | 0.00/1.36M [00:00<?, ?B/s]

config.json:   0%|          | 0.00/718 [00:00<?, ?B/s]

model.safetensors:   0%|          | 0.00/1.52G [00:00<?, ?B/s]

generation_config.json:   0%|          | 0.00/124 [00:00<?, ?B/s]

tokenizer_config.json:   0%|          | 0.00/26.0 [00:00<?, ?B/s]

vocab.json:   0%|          | 0.00/1.04M [00:00<?, ?B/s]

merges.txt:   0%|          | 0.00/456k [00:00<?, ?B/s]

tokenizer.json:   0%|          | 0.00/1.36M [00:00<?, ?B/s]

config.json:   0%|          | 0.00/666 [00:00<?, ?B/s]

model.safetensors:   0%|          | 0.00/3.25G [00:00<?, ?B/s]

generation_config.json:   0%|          | 0.00/124 [00:00<?, ?B/s]

tokenizer_config.json:   0%|          | 0.00/26.0 [00:00<?, ?B/s]

vocab.json:   0%|          | 0.00/1.04M [00:00<?, ?B/s]

merges.txt:   0%|          | 0.00/456k [00:00<?, ?B/s]

tokenizer.json:   0%|          | 0.00/1.36M [00:00<?, ?B/s]
```

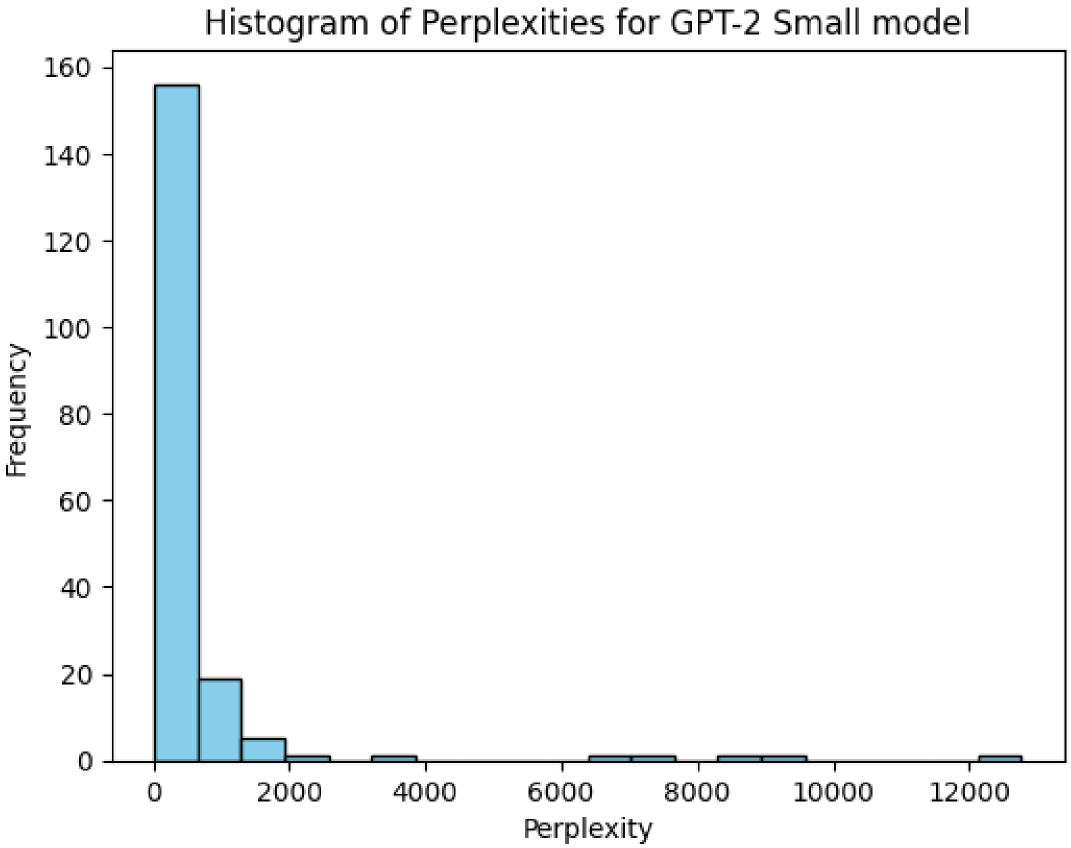
[9]:
```
small_perps = []
medium_perps = []
large_perps = []
for sample in enumerate(tqdm(samples)):
  sample = sample[1]
  small_perps.append(calculate_perplexity(sample, gpt2_small_model,␣
 ↪gpt2_small_tokenizer, device))
  medium_perps.append(calculate_perplexity(sample, gpt2_medium_model,␣
 ↪gpt2_medium_tokenizer, device))
  large_perps.append(calculate_perplexity(sample, gpt2_large_model,␣
 ↪gpt2_large_tokenizer, device))
```

```
100%|     | 187/187 [00:30<00:00,  6.19it/s]
```

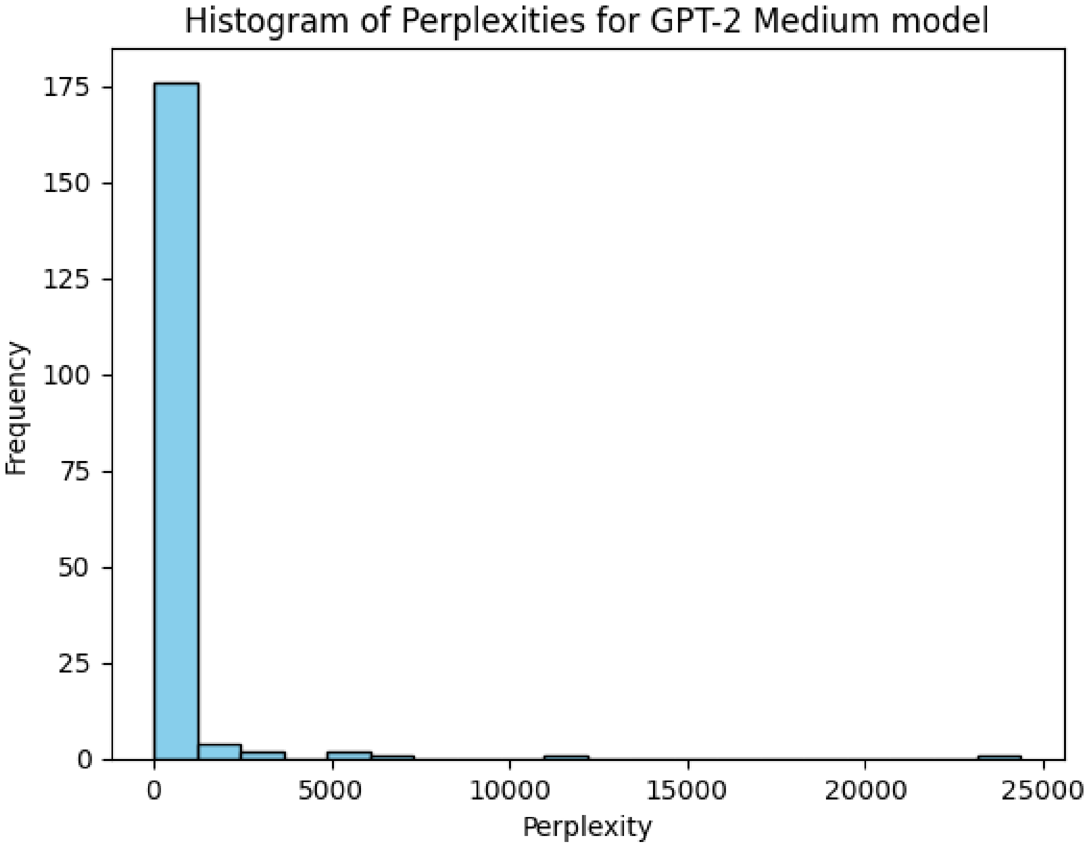
[10]:
```
# Print results:
def plot_perplexities(perplexities, name):
  print(f'Printing results for {name} model')
  # Calculate the 25th and 75th percentiles
  p25 = np.percentile(perplexities, 25)
  p75 = np.percentile(perplexities, 75)
```

```
    # Calculate the midhinge (average of the 25th and 75th percentiles)
    midhinge = (p25 + p75) / 2

    # Calculate the median
    median = np.median(perplexities)

    # Output the values
    print(f"25th Percentile: {p25}")
    print(f"75th Percentile: {p75}")
    print(f"Midhinge: {midhinge}")
    print(f"Median: {median}")

    # Plot a histogram of the perplexities
    plt.hist(perplexities, bins=20, color='skyblue', edgecolor='black')
    plt.title(f'Histogram of Perplexities for {name} model')
    plt.xlabel('Perplexity')
    plt.ylabel('Frequency')
    plt.show()

plot_perplexities(small_perps, 'GPT-2 Small')
plot_perplexities(medium_perps, 'GPT-2 Medium')
plot_perplexities(large_perps, 'GPT-2 Large')
```

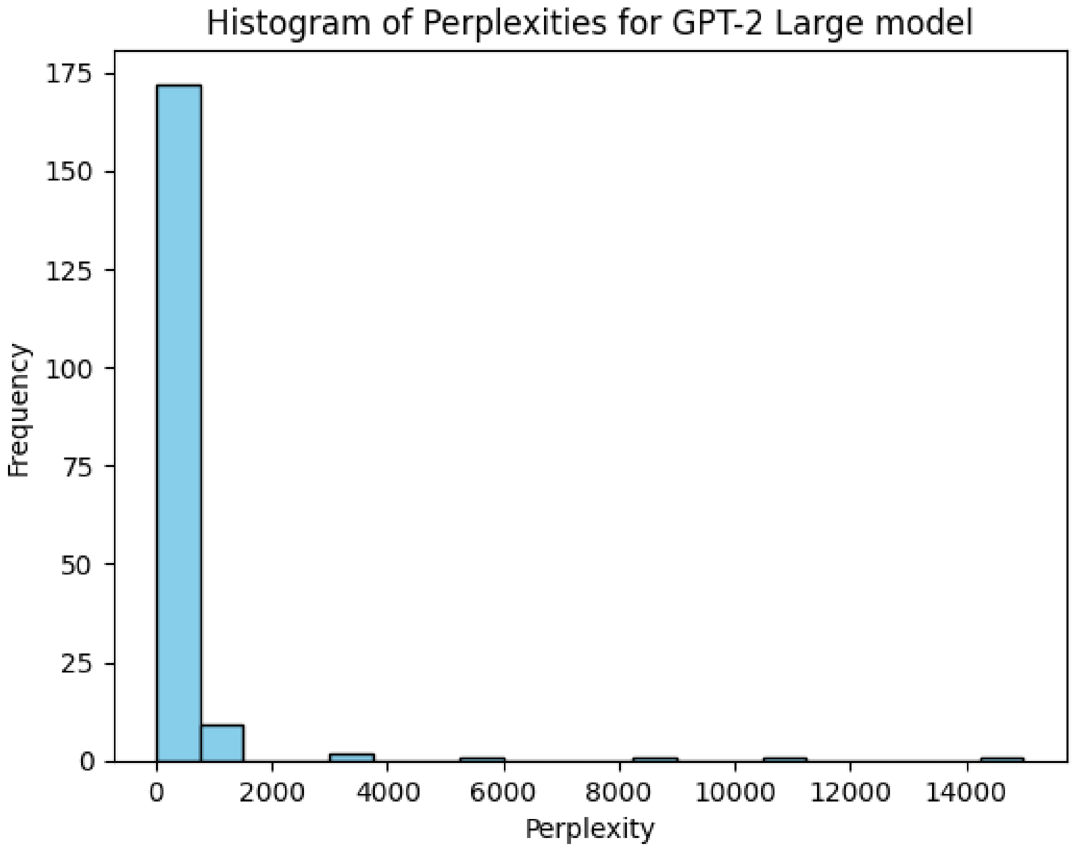
```
Printing results for GPT-2 Small model
25th Percentile: 46.40217567630185
75th Percentile: 276.11094741527165
Midhinge: 161.25656154578675
Median: 67.30752334305188
```



```
Printing results for GPT-2 Medium model
25th Percentile: 34.64231364050474
75th Percentile: 277.53644446766344
Midhinge: 156.08937905408408
Median: 47.35342138334002
```



```
Printing results for GPT-2 Large model
25th Percentile: 29.106181351205585
75th Percentile: 235.48645138789593
Midhinge: 132.29631636955077
Median: 41.22056972785153
```



1.3 c)

The perplexities of all the three models are right skewed meaning the models are not being surprised in generating the next token in the sequence and are performing relatively well.

[14]:
```
# Finding 3 sentences with lowest and highest perplexities.
x = np.argsort(large_perps)

# Printing top 3 sentences with lowest perplexities
print('\t\tTop 3 Sentences with Lowest Perplexities')
for i in range(3):
    print(f'Sample {i+1}')
    print(f'Perplexity: {large_perps[x[i]]}')
    print(f'Sentence: {samples[x[i]]}')
    print()
```

```
                Top 3 Sentences with Lowest Perplexities
Sample 1
Perplexity: 13.815390057070363
Sentence:  The ships ' secondary armament consisted of twenty 50 @-@ calibre 14
@-@ centimetre Type 3 . Eighteen of these were mounted in casemates in the
```

```
forecastle and superstructure and the remaining pair were mounted on the deck
above them and protected by gun shields . They had a maximum elevation of + 20
degrees which gave them ranges of 16 @,@ 300 metres ( 17 @,@ 800 yd ) . Each gun
had a rate of fire of up to 10 rounds per minute . Anti @-@ aircraft defence was
provided by four 40 @-@ calibre 3rd Year Type 8 @-@ centimetre AA guns in single
mounts . The 7 @.@ 62 @-@ centimetre ( 3 in ) high @-@ angle guns had a maximum
elevation of + 75 degrees , and had a rate of fire of 13 to 20 rounds per minute
. They fired a 6 kg ( 13 lb ) projectile with a muzzle velocity of 680 m / s ( 2
@,@ 200 ft / s ) to a maximum height of 7 @,@ 500 metres ( 24 @,@ 600 ft ) . The
ships were also fitted with six submerged 53 @.@ 3 @-@ centimetre ( 21 @.@ 0 in
) torpedo tubes , three on each broadside . They carried twelve to eighteen 6th
Year Type torpedoes which had a 200 @-@ kilogram ( 440 lb ) warhead . They had
three settings for range and speed : 15 @,@ 000 metres ( 16 @,@ 000 yd ) at 26
knots ( 48 km / h ; 30 mph ) , 10 @,@ 000 metres ( 11 @,@ 000 yd ) at 32 knots (
59 km / h ; 37 mph ) , or 7 @,@ 000 metres ( 7 @,@ 700 yd ) at 37 knots ( 69 km
/ h ; 43 mph ) .


Sample 2
Perplexity: 14.7471495641753
Sentence:  The single made its Irish Singles Chart debut at number 24 on the
week ending 13 December 2012 . It peaked at number seven on the week ending 17
January 2013 , marking their sixth top ten appearance in Ireland . " Kiss You "
entered at number 152 in the UK Singles Chart on 24 November 2012 . It peaked at
number nine on the UK Singles Chart on 26 January 2013 , becoming One Direction
's sixth top ten hit in the United Kingdom . On the week ending 18 November 2012
, " Kiss You " debuted at number 90 on the United States Billboard Hot 100 due
to digital download sales from its parent album . As a result of an " end @-@ of
@-@ year download rush " on the week ending 30 December 2012 , the track re @-@
entered the Hot 100 at number 83 . After the accompanying music video was
released , the song re @-@ entered the Hot 100 at number 65 . " Kiss You " had
sold 207 @,@ 000 digital downloads in the US by 18 January 2013 . The single
ultimately peaked at number 46 on the Hot 100 and was certified gold by the
Recording Industry Association of America ( RIAA ) on 25 April 2013 , denoting
shipments of 500 @,@ 000 copies .


Sample 3
Perplexity: 15.953075809850512
Sentence:  The song became One Direction 's fourth top @-@ forty hit on the
Canadian Hot 100 , peaking at number 30 . The single bowed at number 13 on the
Australian Singles Chart on 27 January 2013 , marking its peak position and the
group 's fourth top twenty hit in Australia . The song has been certified
platinum by the Australian Recording Industry Association ( ARIA ) for shipments
of 70 @,@ 000 copies . The track entered the New Zealand Singles Chart at number
17 on 11 January 2013 . It peaked at number 13 in its third and fourth charting
weeks , becominh the group 's sixth top @-@ forty appearance in New Zealand . "
Kiss You " has received a gold certification from the Recording Industry
```

```
Association of New Zealand ( RIANZ ) , indicating sales of 7 @,@ 500 copies .
The track also reached the top 40 in both Belgian territories ( Flanders and
Wallonia ) , as well as in the Czech Republic , Denmark , France , the
Netherlands , and South Korea . In addition , " Kiss You " received gold
certifications from the IFPI Norway and Denmark associations , signifying
collective shipments of 20 @,@ 000 units .
```

[15]:

```python
# Finding 3 sentences with lowest and highest perplexities.
x = np.argsort(large_perps)[::-1]

# Printing top 3 sentences with lowest perplexities
print('\t\tTop 3 Sentences with Highest Perplexities')
for i in range(3):
    print(f'Sample {i+1}')
    print(f'Perplexity: {large_perps[x[i]]}')
    print(f'Sentence: {samples[x[i]]}')
    print()
```

```
                Top 3 Sentences with Highest Perplexities
Sample 1
Perplexity: 14977.117170659934
Sentence:  = 1933 Treasure Coast hurricane =



Sample 2
Perplexity: 10631.543851404731
Sentence:  CD single



Sample 3
Perplexity: 8314.274861806212
Sentence:  = Du Fu =
```

The sentences with high perplexities convey the reason as they are relatively shorter, lack context, and the specific event may be relatively rare or uncommon in the training data, leading to uncertainty for the model.

On the contrary, sentences with lower perplexity values are having detailed description, filled with specific and structured information. Despite their complexity, it has low perplexity because the language and format are highly structured, likely resembling other technical descriptions in the model's training data, making it easier for the model to predict.

2 Problem 2

[11]:

```
# Installing dependencies
!pip install --quiet faiss-gpu mauve-text
```

```
/opt/conda/lib/python3.10/pty.py:89: RuntimeWarning: os.fork() was called.
os.fork() is incompatible with multithreaded code, and JAX is multithreaded, so
this will likely lead to a deadlock.
  pid, fd = os.forkpty()
```

[59]:

```
#Imports
from transformers import GPT2Tokenizer, GPT2LMHeadModel

from tqdm import tqdm
import torch
import json
import os
import sys
import requests
import pickle
import evaluate
import mauve
import numpy as np

seed = 123
torch.manual_seed(seed)
np.random.seed(seed)
```

[ ]:

```
# Load test set
with open('/kaggle/input/data-hw1-twai/Problem2_dataset.pkl', 'rb') as f:
    dataset = pickle.load(f)
```

[16]:

```
print('First 2 test samples:')
for i, example in enumerate(dataset[:2]):
  print(f'{i+1}. {example["text"]}')
```

```
First 2 test samples:
1. Is this restaurant family-friendly ? Yes No Unsure

Does this restaurant accept reservations ? Yes No Unsure

Are the prices at this restaurant mid-range / moderate? Yes No Unsure

Is this restaurant good for dinner? Yes No Unsure

Is this restaurant good for lunch? Yes No Unsure

Is this a Japanese restaurant? Yes No Unsure
```

Could this location be considered a specialty food market? Yes No Unsure

Does this restaurant have waiters and waitresses ? Yes No Unsure

Is this restaurant a hidden gem or off-the-beaten path? Yes No Unsure

Is this primarily a bakery ? Yes No Unsure
2. Clinton talks about her time of 'reflection' during sick days

Hillary Clinton returned to the campaign trail Thursday afternoon, debuting some new intro music and telling the crowd that her sick days allowed her a chance to "reconnect with what this whole campaign is about."

The former secretary of state, who took the stage to James Brown's "I Feel Good," spent the beginning of the week at her home in Chappaqua, New York, after being diagnosed late last week with pneumonia. Her campaign initially did not disclose the illness and only did so after Clinton was forced to leave an event early on Sunday commemorating the 15th anniversary of the Sept. 11 terrorist attacks.

Story Continued Below

"I tried to power through it but even I had to admit that maybe a few days of rest would do me good," she told the Greensboro, North Carolina, crowd of her pneumonia. "And I'm not great at taking it easy, even under ordinary circumstances, but with just two months to go until Election Day, sitting at home was pretty much the last place I wanted to be."

"But it turns out having a few days to myself was actually a gift, I talked with some old friends. I spent time with our very sweet dogs. I did some thinking," she continued. "The campaign trail doesn't really encourage reflection. And it's important to sit with your thoughts every now and then and that did help me reconnect with what this whole campaign is about."

Clinton compared her own ability to take a handful of sick days to that of many Americans who she said are forced to "either go to work sick or they lose a paycheck." She said those Americans, and others "living on a razor's edge" with an aging parent who needs help or without the means to afford a college education, are the reason she is running for president.

Speaking in North Carolina, Clinton made special mention of the law there forcing transgender individuals to use the bathroom that corresponds with the gender on their birth certificate. The law, seen by many as discriminatory, has prompted the NBA, NCAA and Atlantic Coast Conference to move major sporting events out of the state, costs that the former secretary of state said, "We can't afford."

"I'm running for the LGBT teenager here in North Carolina who sees your governor

```
sign a bill legalizing discrimination and suddenly feels like a second-class
citizen," she said. "And if anyone wonders what is the cost of discrimination
are, just ask the people and businesses of North Carolina, look at what's
happening with the NCAA and the ACC. This is where bigotry leads and we can't
afford it, not here, not anywhere else in America."

Clinton did not address her opponent, Donald Trump, by name but did spend a
significant portion of her remarks laying out many of the contrasts she
regularly draws with her opponent. She told the crowd that "I am actually asking
Americans to hold me accountable for my ideas and hold my opponent accountable
for his."

"You know, I've been involved in politics. It is not an easy business. It can
get rough, and I've built up some defenses. When it comes to public service, I
am better at the service part than the public part," she said. "People accuse me
of all kinds of things. You probably have seen that. But nobody ever accuses me
of quitting. And I will never give up. I'll never walk away. No matter how tough
the going gets."
```

2.1 a)

[3]:

```python
# Function to generate text using greedy decoding
def generate_greedy(inputs, max_new_tokens):
    output = model.generate(**inputs, max_new_tokens=max_new_tokens, do_sample=False)
    return tokenizer.decode(output[0], skip_special_tokens=True)

# Function to generate text using top-k sampling
def generate_top_k(inputs, max_new_tokens, k=10):
    output = model.generate(**inputs, max_new_tokens=max_new_tokens, do_sample=True, top_k=k)
    return tokenizer.decode(output[0], skip_special_tokens=True)

# Function to generate text using beam search
def generate_beam_search(inputs, max_new_tokens, num_beams=3):
    output = model.generate(**inputs, max_new_tokens=max_new_tokens, num_beams=num_beams, early_stopping=True)
    return tokenizer.decode(output[0], skip_special_tokens=True)
```

[13]:

```python
# Load GPT-2 large tokenizer and model
tokenizer = GPT2Tokenizer.from_pretrained('gpt2-large')
model = GPT2LMHeadModel.from_pretrained('gpt2-large').to(device)

# Ensure model is in evaluation mode
model.eval()

# Define max length for generation
```

```
max_gen_length = 50  # 50 tokens as per the assignment
prefix_length = 10    # Use the first 10 tokens as the prefix

# Store generated results for analysis
ground_truth = []  # The original sentences from the dataset
generated_greedy = []
generated_top_k = []
generated_beam = []

# Process the dataset, selecting first 10 tokens as prefix and generating text
for sentence in tqdm(dataset):
    sentence = sentence['text']
    # Tokenize the first 10 tokens of each sentence
    tokenizer.add_special_tokens({'pad_token': '[PAD]'})
    inputs = tokenizer(sentence, return_tensors='pt', padding='max_length',␣
 ↪truncation=True, max_length=prefix_length).to(device)

    # Generate text with different strategies
    generated_greedy.append(generate_greedy(inputs,␣
 ↪max_new_tokens=max_gen_length))
    generated_top_k.append(generate_top_k(inputs,␣
 ↪max_new_tokens=max_gen_length))
    generated_beam.append(generate_beam_search(inputs,␣
 ↪max_new_tokens=max_gen_length))

    # Store the original ground truth sentence
    enc_sentence = tokenizer(sentence, return_tensors='pt', truncation=True,␣
 ↪max_length=50)
    dec_sentence = tokenizer.decode(enc_sentence['input_ids'][0])
    ground_truth.append(dec_sentence)
```

```
/opt/conda/lib/python3.10/site-
packages/transformers/tokenization_utils_base.py:1601: FutureWarning:
`clean_up_tokenization_spaces` was not set. It will be set to `True` by default.
This behavior will be depracted in transformers v4.45, and will be then set to
`False` by default. For more details check this issue:
https://github.com/huggingface/transformers/issues/31884
  warnings.warn(
100%|      | 200/200 [00:01<00:00, 103.31it/s]
```

2.2 b)

[15]:

```
def compute_mauve(generated_text, reference_text):
    return mauve.compute_mauve(p_text=reference_text, q_text=generated_text,␣
 ↪device_id=0, verbose=False)
```

[16]:
```python
mauve_ = {}
mauve_['greedy'] = compute_mauve(generated_greedy, ground_truth).mauve
mauve_['top_k'] = compute_mauve(generated_top_k, ground_truth).mauve
mauve_['beam'] = compute_mauve(generated_beam, ground_truth).mauve

for e in mauve_:
    print(f'Mauve score for {e} method: {mauve_[e]}')
```

```
Featurizing p:   0%|          | 0/200 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/200 [00:00<?, ?it/s]

WARNING clustering 400 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/200 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/200 [00:00<?, ?it/s]

WARNING clustering 400 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/200 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/200 [00:00<?, ?it/s]

WARNING clustering 400 points to 20 centroids: please provide at least 780
training points

Mauve score for greedy method: 0.8229049841900652
Mauve score for top_k method: 0.9035953835349086
Mauve score for beam method: 0.6785862391135051
```

The mauve scores for the three generation strategies are mentioned above with the mauve score for the top_k method being the highest meaning the method is performing better than the other two. The lower mauve score indicates the machine generated text distribution being farther from the ground truth text distribution.

2.3 c)

[12]:
```python
# Compute MAUVE scores
greedy_scores = []
for p, q in zip(generated_greedy, ground_truth):
  greedy_scores.append(compute_mauve(p, q).mauve)
```

```
Featurizing p:   0%|          | 0/228 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/253 [00:00<?, ?it/s]

WARNING clustering 481 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/249 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/280 [00:00<?, ?it/s]
```

```
WARNING clustering 529 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/305 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/276 [00:00<?, ?it/s]

WARNING clustering 581 points to 28 centroids: please provide at least 1092
training points

Featurizing p:   0%|          | 0/193 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/243 [00:00<?, ?it/s]

WARNING clustering 436 points to 19 centroids: please provide at least 741
training points

Featurizing p:   0%|          | 0/150 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/203 [00:00<?, ?it/s]

WARNING clustering 353 points to 15 centroids: please provide at least 585
training points

Featurizing p:   0%|          | 0/212 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/260 [00:00<?, ?it/s]

WARNING clustering 472 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/258 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/288 [00:00<?, ?it/s]

WARNING clustering 546 points to 26 centroids: please provide at least 1014
training points

Featurizing p:   0%|          | 0/232 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/232 [00:00<?, ?it/s]

WARNING clustering 464 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/168 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/215 [00:00<?, ?it/s]

WARNING clustering 383 points to 17 centroids: please provide at least 663
training points

Featurizing p:   0%|          | 0/223 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/285 [00:00<?, ?it/s]

WARNING clustering 508 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/245 [00:00<?, ?it/s]
Featurizing q:   0%|          | 0/269 [00:00<?, ?it/s]

WARNING clustering 514 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/202 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/242 [00:00<?, ?it/s]

WARNING clustering 444 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/236 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/270 [00:00<?, ?it/s]

WARNING clustering 506 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/257 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/316 [00:00<?, ?it/s]

WARNING clustering 573 points to 26 centroids: please provide at least 1014
training points

Featurizing p:   0%|          | 0/227 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/264 [00:00<?, ?it/s]

WARNING clustering 491 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/236 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/285 [00:00<?, ?it/s]

WARNING clustering 521 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/246 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/244 [00:00<?, ?it/s]

WARNING clustering 490 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/206 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/248 [00:00<?, ?it/s]

WARNING clustering 454 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/190 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/251 [00:00<?, ?it/s]

WARNING clustering 441 points to 19 centroids: please provide at least 741
training points
```

```
Featurizing p:   0%|          | 0/237 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/210 [00:00<?, ?it/s]

WARNING clustering 447 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/271 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/187 [00:00<?, ?it/s]

WARNING clustering 458 points to 19 centroids: please provide at least 741
training points

Featurizing p:   0%|          | 0/237 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/252 [00:00<?, ?it/s]

WARNING clustering 489 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/244 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/299 [00:00<?, ?it/s]

WARNING clustering 543 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/299 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/320 [00:00<?, ?it/s]

WARNING clustering 619 points to 30 centroids: please provide at least 1170
training points

Featurizing p:   0%|          | 0/231 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/263 [00:00<?, ?it/s]

WARNING clustering 494 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/222 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/241 [00:00<?, ?it/s]

WARNING clustering 463 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/254 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/263 [00:00<?, ?it/s]

WARNING clustering 517 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/276 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/268 [00:00<?, ?it/s]
```

```
WARNING clustering 544 points to 27 centroids: please provide at least 1053
training points

Featurizing p:   0%|          | 0/241 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/360 [00:00<?, ?it/s]

WARNING clustering 601 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/251 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/272 [00:00<?, ?it/s]

WARNING clustering 523 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/214 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/229 [00:00<?, ?it/s]

WARNING clustering 443 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/216 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/213 [00:00<?, ?it/s]

WARNING clustering 429 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/226 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/250 [00:00<?, ?it/s]

WARNING clustering 476 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/243 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/282 [00:00<?, ?it/s]

WARNING clustering 525 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/227 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/269 [00:00<?, ?it/s]

WARNING clustering 496 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/147 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/185 [00:00<?, ?it/s]

WARNING clustering 332 points to 15 centroids: please provide at least 585
training points

Featurizing p:   0%|          | 0/261 [00:00<?, ?it/s]
Featurizing q:   0%|          | 0/272 [00:00<?, ?it/s]

WARNING clustering 533 points to 26 centroids: please provide at least 1014
training points

Featurizing p:   0%|          | 0/198 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/241 [00:00<?, ?it/s]

WARNING clustering 439 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/238 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/294 [00:00<?, ?it/s]

WARNING clustering 532 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/188 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/255 [00:00<?, ?it/s]

WARNING clustering 443 points to 19 centroids: please provide at least 741
training points

Featurizing p:   0%|          | 0/204 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/242 [00:00<?, ?it/s]

WARNING clustering 446 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/244 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/286 [00:00<?, ?it/s]

WARNING clustering 530 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/296 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/354 [00:00<?, ?it/s]

WARNING clustering 650 points to 30 centroids: please provide at least 1170
training points

Featurizing p:   0%|          | 0/202 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/211 [00:00<?, ?it/s]

WARNING clustering 413 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/250 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/259 [00:00<?, ?it/s]

WARNING clustering 509 points to 25 centroids: please provide at least 975
training points
```

```
Featurizing p:   0%|          | 0/239 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/257 [00:00<?, ?it/s]

WARNING clustering 496 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/189 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/196 [00:00<?, ?it/s]

WARNING clustering 385 points to 19 centroids: please provide at least 741
training points

Featurizing p:   0%|          | 0/237 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/317 [00:00<?, ?it/s]

WARNING clustering 554 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/221 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/298 [00:00<?, ?it/s]

WARNING clustering 519 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/188 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/239 [00:00<?, ?it/s]

WARNING clustering 427 points to 19 centroids: please provide at least 741
training points

Featurizing p:   0%|          | 0/154 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/85 [00:00<?, ?it/s]

WARNING clustering 239 points to 8 centroids: please provide at least 312
training points

Featurizing p:   0%|          | 0/277 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/291 [00:00<?, ?it/s]

WARNING clustering 568 points to 28 centroids: please provide at least 1092
training points

Featurizing p:   0%|          | 0/174 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/50 [00:00<?, ?it/s]

Featurizing p:   0%|          | 0/248 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/242 [00:00<?, ?it/s]

WARNING clustering 490 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/205 [00:00<?, ?it/s]
Featurizing q:   0%|          | 0/260 [00:00<?, ?it/s]

WARNING clustering 465 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/221 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/264 [00:00<?, ?it/s]

WARNING clustering 485 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/238 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/272 [00:00<?, ?it/s]

WARNING clustering 510 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/258 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/263 [00:00<?, ?it/s]

WARNING clustering 521 points to 26 centroids: please provide at least 1014
training points

Featurizing p:   0%|          | 0/184 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/213 [00:00<?, ?it/s]

WARNING clustering 397 points to 18 centroids: please provide at least 702
training points

Featurizing p:   0%|          | 0/261 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/254 [00:00<?, ?it/s]

WARNING clustering 515 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/197 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/183 [00:00<?, ?it/s]

WARNING clustering 380 points to 18 centroids: please provide at least 702
training points

Featurizing p:   0%|          | 0/236 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/255 [00:00<?, ?it/s]

WARNING clustering 491 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/201 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/255 [00:00<?, ?it/s]

WARNING clustering 456 points to 20 centroids: please provide at least 780
training points
```

```
Featurizing p:   0%|          | 0/250 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/266 [00:00<?, ?it/s]

WARNING clustering 516 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/138 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/177 [00:00<?, ?it/s]

WARNING clustering 315 points to 14 centroids: please provide at least 546
training points

Featurizing p:   0%|          | 0/170 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/252 [00:00<?, ?it/s]

WARNING clustering 422 points to 17 centroids: please provide at least 663
training points

Featurizing p:   0%|          | 0/179 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/262 [00:00<?, ?it/s]

WARNING clustering 441 points to 18 centroids: please provide at least 702
training points

Featurizing p:   0%|          | 0/272 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/306 [00:00<?, ?it/s]

WARNING clustering 578 points to 27 centroids: please provide at least 1053
training points

Featurizing p:   0%|          | 0/225 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/235 [00:00<?, ?it/s]

WARNING clustering 460 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/205 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/225 [00:00<?, ?it/s]

WARNING clustering 430 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/128 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/106 [00:00<?, ?it/s]

WARNING clustering 234 points to 11 centroids: please provide at least 429
training points

Featurizing p:   0%|          | 0/231 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/192 [00:00<?, ?it/s]
```

```
WARNING clustering 423 points to 19 centroids: please provide at least 741
training points

Featurizing p:   0%|          | 0/254 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/252 [00:00<?, ?it/s]

WARNING clustering 506 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/233 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/253 [00:00<?, ?it/s]

WARNING clustering 486 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/231 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/234 [00:00<?, ?it/s]

WARNING clustering 465 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/250 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/275 [00:00<?, ?it/s]

WARNING clustering 525 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/210 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/231 [00:00<?, ?it/s]

WARNING clustering 441 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/238 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/245 [00:00<?, ?it/s]

WARNING clustering 483 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/222 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/263 [00:00<?, ?it/s]

WARNING clustering 485 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/227 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/252 [00:00<?, ?it/s]

WARNING clustering 479 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/279 [00:00<?, ?it/s]
Featurizing q:   0%|          | 0/301 [00:00<?, ?it/s]

WARNING clustering 580 points to 28 centroids: please provide at least 1092
training points

Featurizing p:   0%|          | 0/237 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/242 [00:00<?, ?it/s]

WARNING clustering 479 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/168 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/233 [00:00<?, ?it/s]

WARNING clustering 401 points to 17 centroids: please provide at least 663
training points

Featurizing p:   0%|          | 0/259 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/298 [00:00<?, ?it/s]

WARNING clustering 557 points to 26 centroids: please provide at least 1014
training points

Featurizing p:   0%|          | 0/237 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/273 [00:00<?, ?it/s]

WARNING clustering 510 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/222 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/257 [00:00<?, ?it/s]

WARNING clustering 479 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/182 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/253 [00:00<?, ?it/s]

WARNING clustering 435 points to 18 centroids: please provide at least 702
training points

Featurizing p:   0%|          | 0/269 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/266 [00:00<?, ?it/s]

WARNING clustering 535 points to 27 centroids: please provide at least 1053
training points

Featurizing p:   0%|          | 0/232 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/218 [00:00<?, ?it/s]

WARNING clustering 450 points to 22 centroids: please provide at least 858
training points
```

```
Featurizing p:   0%|          | 0/195 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/287 [00:00<?, ?it/s]

WARNING clustering 482 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/200 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/232 [00:00<?, ?it/s]

WARNING clustering 432 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/171 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/222 [00:00<?, ?it/s]

WARNING clustering 393 points to 17 centroids: please provide at least 663
training points

Featurizing p:   0%|          | 0/254 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/270 [00:00<?, ?it/s]

WARNING clustering 524 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/251 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/247 [00:00<?, ?it/s]

WARNING clustering 498 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/199 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/237 [00:00<?, ?it/s]

WARNING clustering 436 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/235 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/251 [00:00<?, ?it/s]

WARNING clustering 486 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/175 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/259 [00:00<?, ?it/s]

WARNING clustering 434 points to 18 centroids: please provide at least 702
training points

Featurizing p:   0%|          | 0/211 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/209 [00:00<?, ?it/s]
```

```
WARNING clustering 420 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/195 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/268 [00:00<?, ?it/s]

WARNING clustering 463 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/221 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/270 [00:00<?, ?it/s]

WARNING clustering 491 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/245 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/264 [00:00<?, ?it/s]

WARNING clustering 509 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/246 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/255 [00:00<?, ?it/s]

WARNING clustering 501 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/191 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/254 [00:00<?, ?it/s]

WARNING clustering 445 points to 19 centroids: please provide at least 741
training points

Featurizing p:   0%|          | 0/226 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/206 [00:00<?, ?it/s]

WARNING clustering 432 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/227 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/237 [00:00<?, ?it/s]

WARNING clustering 464 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/265 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/286 [00:00<?, ?it/s]

WARNING clustering 551 points to 26 centroids: please provide at least 1014
training points

Featurizing p:   0%|          | 0/186 [00:00<?, ?it/s]
Featurizing q:   0%|          | 0/230 [00:00<?, ?it/s]

WARNING clustering 416 points to 19 centroids: please provide at least 741
training points

Featurizing p:   0%|          | 0/171 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/214 [00:00<?, ?it/s]

WARNING clustering 385 points to 17 centroids: please provide at least 663
training points

Featurizing p:   0%|          | 0/260 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/255 [00:00<?, ?it/s]

WARNING clustering 515 points to 26 centroids: please provide at least 1014
training points

Featurizing p:   0%|          | 0/202 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/207 [00:00<?, ?it/s]

WARNING clustering 409 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/235 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/233 [00:00<?, ?it/s]

WARNING clustering 468 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/223 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/229 [00:00<?, ?it/s]

WARNING clustering 452 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/245 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/287 [00:00<?, ?it/s]

WARNING clustering 532 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/189 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/169 [00:00<?, ?it/s]

WARNING clustering 358 points to 17 centroids: please provide at least 663
training points

Featurizing p:   0%|          | 0/257 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/303 [00:00<?, ?it/s]

WARNING clustering 560 points to 26 centroids: please provide at least 1014
training points
```

```
Featurizing p:   0%|          | 0/248 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/278 [00:00<?, ?it/s]

WARNING clustering 526 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/235 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/262 [00:00<?, ?it/s]

WARNING clustering 497 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/199 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/234 [00:00<?, ?it/s]

WARNING clustering 433 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/243 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/226 [00:00<?, ?it/s]

WARNING clustering 469 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/243 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/238 [00:00<?, ?it/s]

WARNING clustering 481 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/243 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/203 [00:00<?, ?it/s]

WARNING clustering 446 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/260 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/277 [00:00<?, ?it/s]

WARNING clustering 537 points to 26 centroids: please provide at least 1014
training points

Featurizing p:   0%|          | 0/175 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/241 [00:00<?, ?it/s]

WARNING clustering 416 points to 18 centroids: please provide at least 702
training points

Featurizing p:   0%|          | 0/251 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/285 [00:00<?, ?it/s]
```

```
WARNING clustering 536 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/213 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/228 [00:00<?, ?it/s]

WARNING clustering 441 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/217 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/130 [00:00<?, ?it/s]

WARNING clustering 347 points to 13 centroids: please provide at least 507
training points

Featurizing p:   0%|          | 0/222 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/205 [00:00<?, ?it/s]

WARNING clustering 427 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/234 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/297 [00:00<?, ?it/s]

WARNING clustering 531 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/207 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/267 [00:00<?, ?it/s]

WARNING clustering 474 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/221 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/281 [00:00<?, ?it/s]

WARNING clustering 502 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/259 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/321 [00:00<?, ?it/s]

WARNING clustering 580 points to 26 centroids: please provide at least 1014
training points

Featurizing p:   0%|          | 0/235 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/253 [00:00<?, ?it/s]

WARNING clustering 488 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/212 [00:00<?, ?it/s]
Featurizing q:   0%|          | 0/246 [00:00<?, ?it/s]

WARNING clustering 458 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/304 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/323 [00:00<?, ?it/s]

WARNING clustering 627 points to 30 centroids: please provide at least 1170
training points

Featurizing p:   0%|          | 0/224 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/247 [00:00<?, ?it/s]

WARNING clustering 471 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/207 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/211 [00:00<?, ?it/s]

WARNING clustering 418 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/294 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/307 [00:00<?, ?it/s]

WARNING clustering 601 points to 29 centroids: please provide at least 1131
training points

Featurizing p:   0%|          | 0/163 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/221 [00:00<?, ?it/s]

WARNING clustering 384 points to 16 centroids: please provide at least 624
training points

Featurizing p:   0%|          | 0/233 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/253 [00:00<?, ?it/s]

WARNING clustering 486 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/242 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/260 [00:00<?, ?it/s]

WARNING clustering 502 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/245 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/276 [00:00<?, ?it/s]

WARNING clustering 521 points to 24 centroids: please provide at least 936
training points
```

```
Featurizing p:   0%|          | 0/220 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/142 [00:00<?, ?it/s]

WARNING clustering 362 points to 14 centroids: please provide at least 546
training points

Featurizing p:   0%|          | 0/221 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/234 [00:00<?, ?it/s]

WARNING clustering 455 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/263 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/282 [00:00<?, ?it/s]

WARNING clustering 545 points to 26 centroids: please provide at least 1014
training points

Featurizing p:   0%|          | 0/204 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/224 [00:00<?, ?it/s]

WARNING clustering 428 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/228 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/262 [00:00<?, ?it/s]

WARNING clustering 490 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/174 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/139 [00:00<?, ?it/s]

WARNING clustering 313 points to 14 centroids: please provide at least 546
training points

Featurizing p:   0%|          | 0/186 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/220 [00:00<?, ?it/s]

WARNING clustering 406 points to 19 centroids: please provide at least 741
training points

Featurizing p:   0%|          | 0/143 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/182 [00:00<?, ?it/s]

WARNING clustering 325 points to 14 centroids: please provide at least 546
training points

Featurizing p:   0%|          | 0/235 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/256 [00:00<?, ?it/s]
```

```
WARNING clustering 491 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/254 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/290 [00:00<?, ?it/s]

WARNING clustering 544 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/155 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/264 [00:00<?, ?it/s]

WARNING clustering 419 points to 16 centroids: please provide at least 624
training points

Featurizing p:   0%|          | 0/220 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/281 [00:00<?, ?it/s]

WARNING clustering 501 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/239 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/277 [00:00<?, ?it/s]

WARNING clustering 516 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/184 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/230 [00:00<?, ?it/s]

WARNING clustering 414 points to 18 centroids: please provide at least 702
training points

Featurizing p:   0%|          | 0/208 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/279 [00:00<?, ?it/s]

WARNING clustering 487 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/294 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/298 [00:00<?, ?it/s]

WARNING clustering 592 points to 29 centroids: please provide at least 1131
training points

Featurizing p:   0%|          | 0/180 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/178 [00:00<?, ?it/s]

WARNING clustering 358 points to 18 centroids: please provide at least 702
training points

Featurizing p:   0%|          | 0/227 [00:00<?, ?it/s]
Featurizing q:   0%|          | 0/256 [00:00<?, ?it/s]

WARNING clustering 483 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/214 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/276 [00:00<?, ?it/s]

WARNING clustering 490 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/258 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/231 [00:00<?, ?it/s]

WARNING clustering 489 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/214 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/287 [00:00<?, ?it/s]

WARNING clustering 501 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/203 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/213 [00:00<?, ?it/s]

WARNING clustering 416 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/236 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/238 [00:00<?, ?it/s]

WARNING clustering 474 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/228 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/236 [00:00<?, ?it/s]

WARNING clustering 464 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/175 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/235 [00:00<?, ?it/s]

WARNING clustering 410 points to 18 centroids: please provide at least 702
training points

Featurizing p:   0%|          | 0/259 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/252 [00:00<?, ?it/s]

WARNING clustering 511 points to 25 centroids: please provide at least 975
training points
```

```
Featurizing p:   0%|          | 0/211 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/243 [00:00<?, ?it/s]

WARNING clustering 454 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/210 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/290 [00:00<?, ?it/s]

WARNING clustering 500 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/254 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/297 [00:00<?, ?it/s]

WARNING clustering 551 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/206 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/160 [00:00<?, ?it/s]

WARNING clustering 366 points to 16 centroids: please provide at least 624
training points

Featurizing p:   0%|          | 0/236 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/231 [00:00<?, ?it/s]

WARNING clustering 467 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/240 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/260 [00:00<?, ?it/s]

WARNING clustering 500 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/261 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/269 [00:00<?, ?it/s]

WARNING clustering 530 points to 26 centroids: please provide at least 1014
training points

Featurizing p:   0%|          | 0/228 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/244 [00:00<?, ?it/s]

WARNING clustering 472 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/174 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/265 [00:00<?, ?it/s]
```

```
WARNING clustering 439 points to 17 centroids: please provide at least 663
training points

Featurizing p:   0%|          | 0/217 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/273 [00:00<?, ?it/s]

WARNING clustering 490 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/210 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/241 [00:00<?, ?it/s]

WARNING clustering 451 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/276 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/239 [00:00<?, ?it/s]

WARNING clustering 515 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/203 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/263 [00:00<?, ?it/s]

WARNING clustering 466 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/200 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/142 [00:00<?, ?it/s]

WARNING clustering 342 points to 14 centroids: please provide at least 546
training points

Featurizing p:   0%|          | 0/243 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/276 [00:00<?, ?it/s]

WARNING clustering 519 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/227 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/268 [00:00<?, ?it/s]

WARNING clustering 495 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/249 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/308 [00:00<?, ?it/s]

WARNING clustering 557 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/241 [00:00<?, ?it/s]
Featurizing q:   0%|          | 0/308 [00:00<?, ?it/s]

WARNING clustering 549 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/219 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/247 [00:00<?, ?it/s]

WARNING clustering 466 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/232 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/262 [00:00<?, ?it/s]

WARNING clustering 494 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/208 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/240 [00:00<?, ?it/s]

WARNING clustering 448 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/241 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/221 [00:00<?, ?it/s]

WARNING clustering 462 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/241 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/213 [00:00<?, ?it/s]

WARNING clustering 454 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/199 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/270 [00:00<?, ?it/s]

WARNING clustering 469 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/228 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/247 [00:00<?, ?it/s]

WARNING clustering 475 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/208 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/234 [00:00<?, ?it/s]

WARNING clustering 442 points to 21 centroids: please provide at least 819
training points
```

```
Featurizing p:   0%|          | 0/211 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/239 [00:00<?, ?it/s]

WARNING clustering 450 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/211 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/254 [00:00<?, ?it/s]

WARNING clustering 465 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/245 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/268 [00:00<?, ?it/s]

WARNING clustering 513 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/237 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/284 [00:00<?, ?it/s]

WARNING clustering 521 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/212 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/300 [00:00<?, ?it/s]

WARNING clustering 512 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/273 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/287 [00:00<?, ?it/s]

WARNING clustering 560 points to 27 centroids: please provide at least 1053
training points

Featurizing p:   0%|          | 0/221 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/224 [00:00<?, ?it/s]

WARNING clustering 445 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/200 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/200 [00:00<?, ?it/s]

WARNING clustering 400 points to 20 centroids: please provide at least 780
training points
```

[14]:

```python
top_k_scores = []
for p, q in zip(generated_top_k, ground_truth):
  top_k_scores.append(compute_mauve(p, q).mauve)
```

```
Featurizing p:   0%|          | 0/207 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/180 [00:00<?, ?it/s]

WARNING clustering 387 points to 18 centroids: please provide at least 702
training points

Featurizing p:   0%|          | 0/230 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/227 [00:00<?, ?it/s]

WARNING clustering 457 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/271 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/214 [00:00<?, ?it/s]

WARNING clustering 485 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/293 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/258 [00:00<?, ?it/s]

WARNING clustering 551 points to 26 centroids: please provide at least 1014
training points

Featurizing p:   0%|          | 0/248 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/214 [00:00<?, ?it/s]

WARNING clustering 462 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/214 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/203 [00:00<?, ?it/s]

WARNING clustering 417 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/253 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/236 [00:00<?, ?it/s]

WARNING clustering 489 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/269 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/228 [00:00<?, ?it/s]

WARNING clustering 497 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/249 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/175 [00:00<?, ?it/s]
```

```
WARNING clustering 424 points to 18 centroids: please provide at least 702
training points

Featurizing p:   0%|          | 0/297 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/259 [00:00<?, ?it/s]

WARNING clustering 556 points to 26 centroids: please provide at least 1014
training points

Featurizing p:   0%|          | 0/237 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/211 [00:00<?, ?it/s]

WARNING clustering 448 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/279 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/210 [00:00<?, ?it/s]

WARNING clustering 489 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/332 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/254 [00:00<?, ?it/s]

WARNING clustering 586 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/278 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/206 [00:00<?, ?it/s]

WARNING clustering 484 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/220 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/236 [00:00<?, ?it/s]

WARNING clustering 456 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/258 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/240 [00:00<?, ?it/s]

WARNING clustering 498 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/267 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/261 [00:00<?, ?it/s]

WARNING clustering 528 points to 26 centroids: please provide at least 1014
training points

Featurizing p:   0%|          | 0/278 [00:00<?, ?it/s]
Featurizing q:   0%|          | 0/228 [00:00<?, ?it/s]

WARNING clustering 506 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/248 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/174 [00:00<?, ?it/s]

WARNING clustering 422 points to 17 centroids: please provide at least 663
training points

Featurizing p:   0%|          | 0/266 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/217 [00:00<?, ?it/s]

WARNING clustering 483 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/273 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/210 [00:00<?, ?it/s]

WARNING clustering 483 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/287 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/276 [00:00<?, ?it/s]

WARNING clustering 563 points to 28 centroids: please provide at least 1092
training points

Featurizing p:   0%|          | 0/281 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/203 [00:00<?, ?it/s]

WARNING clustering 484 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/241 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/200 [00:00<?, ?it/s]

WARNING clustering 441 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/247 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/243 [00:00<?, ?it/s]

WARNING clustering 490 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/267 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/227 [00:00<?, ?it/s]

WARNING clustering 494 points to 23 centroids: please provide at least 897
training points
```

```
Featurizing p:   0%|          | 0/290 [00:00<?, ?it/s]
Featurizing q:   0%|          | 0/249 [00:00<?, ?it/s]
WARNING clustering 539 points to 25 centroids: please provide at least 975
training points
Featurizing p:   0%|          | 0/317 [00:00<?, ?it/s]
Featurizing q:   0%|          | 0/241 [00:00<?, ?it/s]
WARNING clustering 558 points to 24 centroids: please provide at least 936
training points
Featurizing p:   0%|          | 0/260 [00:00<?, ?it/s]
Featurizing q:   0%|          | 0/219 [00:00<?, ?it/s]
WARNING clustering 479 points to 22 centroids: please provide at least 858
training points
Featurizing p:   0%|          | 0/262 [00:00<?, ?it/s]
Featurizing q:   0%|          | 0/232 [00:00<?, ?it/s]
WARNING clustering 494 points to 23 centroids: please provide at least 897
training points
Featurizing p:   0%|          | 0/251 [00:00<?, ?it/s]
Featurizing q:   0%|          | 0/208 [00:00<?, ?it/s]
WARNING clustering 459 points to 21 centroids: please provide at least 819
training points
Featurizing p:   0%|          | 0/298 [00:00<?, ?it/s]
Featurizing q:   0%|          | 0/241 [00:00<?, ?it/s]
WARNING clustering 539 points to 24 centroids: please provide at least 936
training points
Featurizing p:   0%|          | 0/256 [00:00<?, ?it/s]
Featurizing q:   0%|          | 0/241 [00:00<?, ?it/s]
WARNING clustering 497 points to 24 centroids: please provide at least 936
training points
Featurizing p:   0%|          | 0/222 [00:00<?, ?it/s]
Featurizing q:   0%|          | 0/199 [00:00<?, ?it/s]
WARNING clustering 421 points to 20 centroids: please provide at least 780
training points
Featurizing p:   0%|          | 0/282 [00:00<?, ?it/s]
Featurizing q:   0%|          | 0/228 [00:00<?, ?it/s]
```

```
WARNING clustering 510 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/220 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/208 [00:00<?, ?it/s]

WARNING clustering 428 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/281 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/211 [00:00<?, ?it/s]

WARNING clustering 492 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/260 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/211 [00:00<?, ?it/s]

WARNING clustering 471 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/239 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/245 [00:00<?, ?it/s]

WARNING clustering 484 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/281 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/237 [00:00<?, ?it/s]

WARNING clustering 518 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/305 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/212 [00:00<?, ?it/s]

WARNING clustering 517 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/322 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/273 [00:00<?, ?it/s]

WARNING clustering 595 points to 27 centroids: please provide at least 1053
training points

Featurizing p:   0%|          | 0/254 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/221 [00:00<?, ?it/s]

WARNING clustering 475 points to 22 centroids: please provide at least 858
training points
```

[ ]:

```
beam_scores = []
for p, q in zip(generated_beam, ground_truth):
  beam_scores.append(compute_mauve(p, q).mauve)
```

```
Featurizing p:   0%|          | 0/268 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/228 [00:00<?, ?it/s]

WARNING clustering 496 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/223 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/249 [00:00<?, ?it/s]

WARNING clustering 472 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/293 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/305 [00:00<?, ?it/s]

WARNING clustering 598 points to 29 centroids: please provide at least 1131
training points

Featurizing p:   0%|          | 0/95 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/193 [00:00<?, ?it/s]

WARNING clustering 288 points to 10 centroids: please provide at least 390
training points

Featurizing p:   0%|          | 0/141 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/150 [00:00<?, ?it/s]

WARNING clustering 291 points to 14 centroids: please provide at least 546
training points

Featurizing p:   0%|          | 0/230 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/212 [00:00<?, ?it/s]

WARNING clustering 442 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/292 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/258 [00:00<?, ?it/s]

WARNING clustering 550 points to 26 centroids: please provide at least 1014
training points

Featurizing p:   0%|          | 0/232 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/232 [00:00<?, ?it/s]

WARNING clustering 464 points to 23 centroids: please provide at least 897
training points
```

```
Featurizing p:   0%|          | 0/222 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/168 [00:00<?, ?it/s]

WARNING clustering 390 points to 17 centroids: please provide at least 663
training points

Featurizing p:   0%|          | 0/273 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/223 [00:00<?, ?it/s]

WARNING clustering 496 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/231 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/245 [00:00<?, ?it/s]

WARNING clustering 476 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/252 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/202 [00:00<?, ?it/s]

WARNING clustering 454 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/272 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/236 [00:00<?, ?it/s]

WARNING clustering 508 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/284 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/257 [00:00<?, ?it/s]

WARNING clustering 541 points to 26 centroids: please provide at least 1014
training points

Featurizing p:   0%|          | 0/272 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/227 [00:00<?, ?it/s]

WARNING clustering 499 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/273 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/236 [00:00<?, ?it/s]

WARNING clustering 509 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/252 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/246 [00:00<?, ?it/s]
```

```
WARNING clustering 498 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/244 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/206 [00:00<?, ?it/s]

WARNING clustering 450 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/213 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/190 [00:00<?, ?it/s]

WARNING clustering 403 points to 19 centroids: please provide at least 741
training points

Featurizing p:   0%|          | 0/231 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/237 [00:00<?, ?it/s]

WARNING clustering 468 points to 23 centroids: please provide at least 897
training points

Featurizing p:   0%|          | 0/185 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/271 [00:00<?, ?it/s]

WARNING clustering 456 points to 18 centroids: please provide at least 702
training points

Featurizing p:   0%|          | 0/185 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/237 [00:00<?, ?it/s]

WARNING clustering 422 points to 18 centroids: please provide at least 702
training points

Featurizing p:   0%|          | 0/289 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/244 [00:00<?, ?it/s]

WARNING clustering 533 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/317 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/299 [00:00<?, ?it/s]

WARNING clustering 616 points to 30 centroids: please provide at least 1170
training points

Featurizing p:   0%|          | 0/193 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/231 [00:00<?, ?it/s]

WARNING clustering 424 points to 19 centroids: please provide at least 741
training points

Featurizing p:   0%|          | 0/253 [00:00<?, ?it/s]
Featurizing q:   0%|          | 0/222 [00:00<?, ?it/s]

WARNING clustering 475 points to 22 centroids: please provide at least 858
training points

Featurizing p:   0%|          | 0/247 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/254 [00:00<?, ?it/s]

WARNING clustering 501 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/252 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/276 [00:00<?, ?it/s]

WARNING clustering 528 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/279 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/241 [00:00<?, ?it/s]

WARNING clustering 520 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/294 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/251 [00:00<?, ?it/s]

WARNING clustering 545 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/235 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/214 [00:00<?, ?it/s]

WARNING clustering 449 points to 21 centroids: please provide at least 819
training points

Featurizing p:   0%|          | 0/186 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/216 [00:00<?, ?it/s]

WARNING clustering 402 points to 19 centroids: please provide at least 741
training points

Featurizing p:   0%|          | 0/238 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/226 [00:00<?, ?it/s]

IOPub message rate exceeded.
The notebook server will temporarily stop sending output
to the client in order to avoid crashing it.
To change this limit, set the config variable
`--NotebookApp.iopub_msg_rate_limit`.

Current values:
NotebookApp.iopub_msg_rate_limit=1000.0 (msgs/sec)
NotebookApp.rate_limit_window=3.0 (secs)

Featurizing q:   0%|          | 0/238 [00:00<?, ?it/s]

WARNING clustering 502 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/255 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/188 [00:00<?, ?it/s]

WARNING clustering 443 points to 19 centroids: please provide at least 741
training points

Featurizing p:   0%|          | 0/249 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/204 [00:00<?, ?it/s]

WARNING clustering 453 points to 20 centroids: please provide at least 780
training points

Featurizing p:   0%|          | 0/285 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/244 [00:00<?, ?it/s]

WARNING clustering 529 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/313 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/296 [00:00<?, ?it/s]

WARNING clustering 609 points to 30 centroids: please provide at least 1170
training points

Featurizing p:   0%|          | 0/175 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/202 [00:00<?, ?it/s]

WARNING clustering 377 points to 18 centroids: please provide at least 702
training points

Featurizing p:   0%|          | 0/248 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/250 [00:00<?, ?it/s]

WARNING clustering 498 points to 25 centroids: please provide at least 975
training points

Featurizing p:   0%|          | 0/237 [00:00<?, ?it/s]

Featurizing q:   0%|          | 0/239 [00:00<?, ?it/s]

WARNING clustering 476 points to 24 centroids: please provide at least 936
training points

Featurizing p:   0%|          | 0/216 [00:00<?, ?it/s]
```

[60]:

```python
sort_indices_greedy = np.argsort(greedy_scores)

# Printing highest and lowest mauve score values.

print('\t\t Mauve Score Statistics for Greedy Strategy\n')
print(f'Lowest Mauve score {greedy_scores[sort_indices_greedy[0]]}')
print(f'Sentence: {generated_greedy[sort_indices_greedy[0]]}')
print()
print(f'Highest Mauve score {greedy_scores[sort_indices_greedy[-1]]}')
print(f'Sentence: {generated_greedy[sort_indices_greedy[-1]]}')
```

```
                 Mauve Score Statistics for Greedy Strategy

Lowest Mauve score 0.2790984339129409
Sentence: /d u·mə/

(d u·mə)

(d u·mə)

(d u·mə)

(d u·mə)




Highest Mauve score 0.997867564334721
Sentence: Get Aston Villa FC updates directly to your inbox Subscribe Thank you
for subscribing We have more newsletters Show me See our privacy notice Could
not subscribe, try again later Invalid Email

Aston Villa are set to make a £5million bid for Manchester City striker Sergio
Aguero.

The Argentine is
```

[61]:

```python
sort_indices_top_k = np.argsort(top_k_scores)

# Printing highest and lowest mauve score values.

print('\t\t Mauve Score Statistics for Top-k Strategy\n')
print(f'Lowest Mauve score {top_k_scores[sort_indices_top_k[0]]}')
print(f'Sentence: {generated_top_k[sort_indices_top_k[0]]}')
print()
print(f'Highest Mauve score {top_k_scores[sort_indices_top_k[-1]]}')
print(f'Sentence: {generated_top_k[sort_indices_top_k[-1]]}')
```

```
                 Mauve Score Statistics for Top-k Strategy
```

```
Lowest Mauve score 0.29117410325695225
Sentence: SEE SIZE CHART

Destroyed Straight Edge:

-1

-1

-1

-1

-1

-1

-1

-1

-1

-1

-1

-1

Highest Mauve score 0.9991030975228555
Sentence: Get Aston Villa FC updates directly to your inbox Subscribe Thank you
for subscribing We have more newsletters Show me See our privacy notice Could
not subscribe, try again later Invalid Email

Aston Villa are lining up a move for Borussia Dortmund star Ilkay Gundogan.

A move for the German
```

[23]:

```python
sort_indices_beam = np.argsort(beam_scores)

# Printing highest and lowest mauve score values.

print('\t\t Mauve Score Statistics for Beam Search Strategy\n')
print(f'Lowest Mauve score {beam_scores[sort_indices_beam[0]]}')
print(f'Sentence: {generated_beam[sort_indices_beam[0]]}')
print()
print(f'Highest Mauve score {beam_scores[sort_indices_beam[-1]]}')
print(f'Sentence: {generated_beam[sort_indices_beam[-1]]}')
```

```
                 Mauve Score Statistics for Beam Search Strategy

Lowest Mauve score 0.49077748156919876
Sentence: /d u·mə/ [d u mə]

d u mən/ [d u mən]

d u mər/ [d u

Highest Mauve score 0.998143357908815
Sentence: Having trouble viewing the video? Try disabling any ad blocking
extensions currently running on your browser.

You are not allowed to watch this stream ='(

SALT LAKE CITY - A Utah man was arrested after police say he shot and killed a
man who was trying to break into his
```

The MAUVE scores demonstrate the model's ability to generate human-like text across different contexts. The highest score corresponds to a sentence with relatively easier and understandable context, indicating the model's proficiency in generating text that closely matches typical human writing patterns. Conversely, the lowest score is associated with text which is specialized and uncommon in natural language. This insight is valuable for understanding the model's capabilities and potential areas for improvement in generating diverse types of text.

3 Problem 3

[31]:
```
!pip install datasets scikit-learn
```

```
/opt/conda/lib/python3.10/pty.py:89: RuntimeWarning: os.fork() was called.
os.fork() is incompatible with multithreaded code, and JAX is multithreaded, so
this will likely lead to a deadlock.
  pid, fd = os.forkpty()
huggingface/tokenizers: The current process just got forked, after parallelism
has already been used. Disabling parallelism to avoid deadlocks…
To disable this warning, you can either:
        - Avoid using `tokenizers` before the fork if possible
        - Explicitly set the environment variable TOKENIZERS_PARALLELISM=(true |
false)

Requirement already satisfied: datasets in /opt/conda/lib/python3.10/site-
packages (2.21.0)
Requirement already satisfied: scikit-learn in /opt/conda/lib/python3.10/site-
packages (1.2.2)
Requirement already satisfied: filelock in /opt/conda/lib/python3.10/site-
packages (from datasets) (3.15.1)
```

```
Requirement already satisfied: numpy>=1.17 in /opt/conda/lib/python3.10/site-
packages (from datasets) (1.26.4)
Requirement already satisfied: pyarrow>=15.0.0 in
/opt/conda/lib/python3.10/site-packages (from datasets) (16.1.0)
Requirement already satisfied: dill<0.3.9,>=0.3.0 in
/opt/conda/lib/python3.10/site-packages (from datasets) (0.3.8)
Requirement already satisfied: pandas in /opt/conda/lib/python3.10/site-packages
(from datasets) (2.2.2)
Requirement already satisfied: requests>=2.32.2 in
/opt/conda/lib/python3.10/site-packages (from datasets) (2.32.3)
Requirement already satisfied: tqdm>=4.66.3 in /opt/conda/lib/python3.10/site-
packages (from datasets) (4.66.4)
Requirement already satisfied: xxhash in /opt/conda/lib/python3.10/site-packages
(from datasets) (3.4.1)
Requirement already satisfied: multiprocess in /opt/conda/lib/python3.10/site-
packages (from datasets) (0.70.16)
Requirement already satisfied: fsspec<=2024.6.1,>=2023.1.0 in
/opt/conda/lib/python3.10/site-packages (from
fsspec[http]<=2024.6.1,>=2023.1.0->datasets) (2024.6.1)
Requirement already satisfied: aiohttp in /opt/conda/lib/python3.10/site-
packages (from datasets) (3.9.5)
Requirement already satisfied: huggingface-hub>=0.21.2 in
/opt/conda/lib/python3.10/site-packages (from datasets) (0.24.6)
Requirement already satisfied: packaging in /opt/conda/lib/python3.10/site-
packages (from datasets) (21.3)
Requirement already satisfied: pyyaml>=5.1 in /opt/conda/lib/python3.10/site-
packages (from datasets) (6.0.2)
Requirement already satisfied: scipy>=1.3.2 in /opt/conda/lib/python3.10/site-
packages (from scikit-learn) (1.14.0)
Requirement already satisfied: joblib>=1.1.1 in /opt/conda/lib/python3.10/site-
packages (from scikit-learn) (1.4.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/opt/conda/lib/python3.10/site-packages (from scikit-learn) (3.5.0)
Requirement already satisfied: aiosignal>=1.1.2 in
/opt/conda/lib/python3.10/site-packages (from aiohttp->datasets) (1.3.1)
Requirement already satisfied: attrs>=17.3.0 in /opt/conda/lib/python3.10/site-
packages (from aiohttp->datasets) (23.2.0)
Requirement already satisfied: frozenlist>=1.1.1 in
/opt/conda/lib/python3.10/site-packages (from aiohttp->datasets) (1.4.1)
Requirement already satisfied: multidict<7.0,>=4.5 in
/opt/conda/lib/python3.10/site-packages (from aiohttp->datasets) (6.0.5)
Requirement already satisfied: yarl<2.0,>=1.0 in /opt/conda/lib/python3.10/site-
packages (from aiohttp->datasets) (1.9.4)
Requirement already satisfied: async-timeout<5.0,>=4.0 in
/opt/conda/lib/python3.10/site-packages (from aiohttp->datasets) (4.0.3)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/opt/conda/lib/python3.10/site-packages (from huggingface-hub>=0.21.2->datasets)
(4.12.2)
```

```
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in
/opt/conda/lib/python3.10/site-packages (from packaging->datasets) (3.1.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/opt/conda/lib/python3.10/site-packages (from requests>=2.32.2->datasets)
(3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /opt/conda/lib/python3.10/site-
packages (from requests>=2.32.2->datasets) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/opt/conda/lib/python3.10/site-packages (from requests>=2.32.2->datasets)
(1.26.18)
Requirement already satisfied: certifi>=2017.4.17 in
/opt/conda/lib/python3.10/site-packages (from requests>=2.32.2->datasets)
(2024.7.4)
Requirement already satisfied: python-dateutil>=2.8.2 in
/opt/conda/lib/python3.10/site-packages (from pandas->datasets) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /opt/conda/lib/python3.10/site-
packages (from pandas->datasets) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in /opt/conda/lib/python3.10/site-
packages (from pandas->datasets) (2024.1)
Requirement already satisfied: six>=1.5 in /opt/conda/lib/python3.10/site-
packages (from python-dateutil>=2.8.2->pandas->datasets) (1.16.0)
```

[32]:

```python
import torch
from transformers import DistilBertTokenizer,␣
↪DistilBertForSequenceClassification, AdamW
from transformers import get_scheduler
from datasets import load_dataset
from sklearn.metrics import accuracy_score, precision_recall_fscore_support
from torch.utils.data import DataLoader
from tqdm.auto import tqdm
from transformers import TrainingArguments, Trainer
from sklearn.metrics import roc_curve, auc
```

[33]:

```python
# Load IMDb dataset and limit to 1000 samples
dataset = load_dataset('imdb')
train_dataset = dataset['train'].shuffle(seed=42).select(range(1000))
test_dataset = dataset['test'].shuffle(seed=42).select(range(1000))  # Select␣
↪100 samples
```

```
Downloading readme:   0%|          | 0.00/7.81k [00:00<?, ?B/s]

Downloading data:   0%|          | 0.00/21.0M [00:00<?, ?B/s]

Downloading data:   0%|          | 0.00/20.5M [00:00<?, ?B/s]

Downloading data:   0%|          | 0.00/42.0M [00:00<?, ?B/s]

Generating train split:   0%|          | 0/25000 [00:00<?, ? examples/s]

Generating test split:   0%|          | 0/25000 [00:00<?, ? examples/s]
Generating unsupervised split:   0%|          | 0/50000 [00:00<?, ? examples/s]
```

[43]:

```python
print('5 Training examples')
for i, example in enumerate(train_dataset['text'][:5]):
  print(f'{i+1}. {example}')
```

```
5 Training examples
1. There is no relation at all between Fortier and Profiler but the fact that
both are police series about violent crimes. Profiler looks crispy, Fortier
looks classic. Profiler plots are quite simple. Fortier's plot are far more
complicated… Fortier looks more like Prime Suspect, if we have to spot
similarities… The main character is weak and weirdo, but have "clairvoyance".
People like to compare, to judge, to evaluate. How about just enjoying? Funny
thing too, people writing Fortier looks American but, on the other hand, arguing
they prefer American series (!!!). Maybe it's the language, or the spirit, but I
think this series is more English than American. By the way, the actors are
really good and funny. The acting is not superficial at all…
2. This movie is a great. The plot is very true to the book which is a classic
written by Mark Twain. The movie starts of with a scene where Hank sings a song
with a bunch of kids called "when you stub your toe on the moon" It reminds me
of Sinatra's song High Hopes, it is fun and inspirational. The Music is great
throughout and my favorite song is sung by the King, Hank (bing Crosby) and Sir
"Saggy" Sagamore. OVerall a great family movie or even a great Date movie. This
is a movie you can watch over and over again. The princess played by Rhonda
Fleming is gorgeous. I love this movie!! If you liked Danny Kaye in the Court
Jester then you will definitely like this movie.
3. George P. Cosmatos' "Rambo: First Blood Part II" is pure wish-fulfillment.
The United States clearly didn't win the war in Vietnam. They caused damage to
this country beyond the imaginable and this movie continues the fairy story of
the oh-so innocent soldiers. The only bad guys were the leaders of the nation,
who made this war happen. The character of Rambo is perfect to notice this. He
is extremely patriotic, bemoans that US-Americans didn't appreciate and
celebrate the achievements of the single soldier, but has nothing but distrust
for leading officers and politicians. Like every film that defends the war (e.g.
"We Were Soldiers") also this one avoids the need to give a comprehensible
reason for the engagement in South Asia. And for that matter also the reason for
every single US-American soldier that was there. Instead, Rambo gets to take
revenge for the wounds of a whole nation. It would have been better to work on
how to deal with the memories, rather than suppressing them. "Do we get to win
this time?" Yes, you do.
4. In the process of trying to establish the audiences' empathy with Jake Roedel
(Tobey Maguire) the filmmakers slander the North and the Jayhawkers. Missouri
never withdrew from the Union and the Union Army was not an invading force. The
Southerners fought for State's Rights: the right to own slaves, elect crooked
legislatures and judges, and employ a political spoils system. There's nothing
noble in that. The Missourians could have easily traveled east and joined the
Confederate Army.<br /><br />It seems to me that the story has nothing to do
with ambiguity. When Jake leaves the Bushwhackers, it's not because he saw error
```

in his way, he certainly doesn't give himself over to the virtue of the cause of abolition.

5. Yeh, I know -- you're quivering with excitement. Well, *The Secret Lives of Dentists* will not upset your expectations: it's solidly made but essentially unimaginative, truthful but dull. It concerns the story of a married couple who happen to be dentists and who share the same practice (already a recipe for trouble: if it wasn't for our separate work-lives, we'd all ditch our spouses out of sheer irritation). Campbell Scott, whose mustache and demeanor don't recall Everyman so much as Ned Flanders from *The Simpsons*, is the mild-mannered, uber-Dad husband, and Hope Davis is the bored-stiff housewife who channels her frustrations into amateur opera. One night, as Dad & the daughters attend one of Davis' performances, he discovers that his wife is channeling her frustrations into more than just singing: he witnesses his wife kissing and flirting with the director of opera. (One nice touch: we never see the opera-director's face.) Dreading the prospect of instituting the proceedings for separation, divorce, and custody hearings -- profitable only to the lawyers -- Scott chooses to pretend ignorance of his wife's indiscretions.<br /><br />Already, the literate among you are starting to yawn: ho-hum, another story about the Pathetic, Sniveling Little Cuckold. But Rudolph, who took the story from a Jane Smiley novella, hopes that the wellworn-ness of the material will be compensated for by a series of flashy, postmodern touches. For instance, one of Scott's belligerent patients (Denis Leary, kept relatively -- and blessedly -- in check) will later become a sort of construction of the dentist's imagination, emerging as a Devil-on-the-shoulder advocate for the old-fashioned masculine virtues ("Dump the b---h!", etc.). When not egged-on by his imaginary new buddy, Scott is otherwise tormented by fantasies that include his wife engaged in a three-way with two of the male dental-assistants who work in their practice. It's not going too far to say that this movie is *Eyes Wide Shut* for Real People (or Grown-Ups, at least). Along those lines, Campbell Scott and Hope Davis are certainly recognizable human beings as compared to the glamourpuss pair of Cruise and Kidman. Further, the script for *Secret Lives* is clearly more relevant than Kubrick's. As proof, I offer the depiction of the dentists' children, particularly the youngest one who is about 3 or 4 years old, and whose main utterance is "Dad! Dad! Dad! Dad! Dad! DAD!!!" This is Family Life, all right, with all its charms.<br /><br />The movie would make an interesting double-bill with *Kramer vs. Kramer*, as well. One can easily trace the Feminization of the American Male from 1979 to 2003. In this movie, Dad is the housewife as in *Kramer*, but he is in no way flustered by the domestic role, unlike Dustin Hoffman, who was too manly to make toast. Here, Scott gets all the plumb chores, such as wiping up the children's vomit, cooking, cleaning, taking the kids to whatever inane after-school activity is on the docket. And all without complaint. (And without directorial commentary. It's just taken for granted.)<br /><br />The film has virtues, mostly having to do with verisimilitude. However, it's dragged down from greatness by its insistence on trendy distractions, which culminate in a long scene where a horrible five-day stomach flu makes the rounds in the household. We must endure pointless fantasy sequences, initiated by the imaginary ringleader Leary. Whose existence, by the way, is finally reminiscent of the Brad Pitt character in *Fight Club*. And this

```
finally drives home the film's other big flaw: lack of originality. In this
review, I realize it's been far too easy to reference many other films. Granted,
this film is an improvement on most of them, but still. *The Secret Lives of
Dentists* is worth seeing, but don't get too excited about it. (Not that you
were all that excited, anyway. I guess.)
```

[44]:

```
print('5 Testing examples')
for i, example in enumerate(test_dataset['text'][:5]):
  print(f'{i+1}. {example}')
```

```
5 Testing examples
1. <br /><br />When I unsuspectedly rented A Thousand Acres, I thought I was in
for an entertaining King Lear story and of course Michelle Pfeiffer was in it,
so what could go wrong?<br /><br />Very quickly, however, I realized that this
story was about A Thousand Other Things besides just Acres. I started crying and
couldn't stop until long after the movie ended. Thank you Jane, Laura and
Jocelyn, for bringing us such a wonderfully subtle and compassionate movie!
Thank you cast, for being involved and portraying the characters with such depth
and gentleness!<br /><br />I recognized the Angry sister; the Runaway sister and
the sister in Denial. I recognized the Abusive Husband and why he was there and
then the Father, oh oh the Father… all superbly played. I also recognized
myself and this movie was an eye-opener, a relief, a chance to face my OWN truth
and finally doing something about it. I truly hope A Thousand Acres has had the
same effect on some others out there.<br /><br />Since I didn't understand why
the cover said the film was about sisters fighting over land -they weren't
fighting each other at all- I watched it a second time. Then I was able to see
that if one hadn't lived a similar story, one would easily miss the overwhelming
undercurrent of dread and fear and the deep bond between the sisters that runs
through it all. That is exactly the reason why people in general often overlook
the truth about their neighbors for instance.<br /><br />But yet another reason
why this movie is so perfect!<br /><br />I don't give a rat's ass (pardon my
French) about to what extend the King Lear story is followed. All I know is that
I can honestly say: this movie has changed my life.<br /><br />Keep up the good
work guys, you CAN and DO make a difference.<br /><br />
2. This is the latest entry in the long series of films with the French agent,
O.S.S. 117 (the French answer to James Bond). The series was launched in the
early 1950's, and spawned at least eight films (none of which was ever released
in the U.S.). 'O.S.S.117:Cairo,Nest Of Spies' is a breezy little comedy that
should not…repeat NOT, be taken too seriously. Our protagonist finds himself
in the middle of a spy chase in Egypt (with Morroco doing stand in for Egypt) to
find out about a long lost friend. What follows is the standard James
Bond/Inspector Cloussou kind of antics. Although our man is something of an
overt xenophobe,sexist,homophobe, it's treated as pure farce (as I said, don't
take it too seriously). Although there is a bit of rough language & cartoon
violence, it's basically okay for older kids (ages 12 & up). As previously
stated in the subject line, just sit back,pass the popcorn & just enjoy.
3. This movie was so frustrating. Everything seemed energetic and I was totally
prepared to have a good time. I at least thought I'd be able to stand it. But, I
```

```
was wrong. First, the weird looping? It was like watching "America's Funniest
Home Videos". The damn parents. I hated them so much. The stereo-typical Latino
family? I need to speak with the person responsible for this. We need to have a
talk. That little girl who was always hanging on someone? I just hated her and
had to mention it. Now, the final scene transcends, I must say. It's so
gloriously bad and full of badness that it is a movie of its own. What crappy
dancing. Horrible and beautiful at once.
4. I was truly and wonderfully surprised at "O' Brother, Where Art Thou?" The
video store was out of all the movies I was planning on renting, so then I came
across this. I came home and as I watched I became engrossed and found myself
laughing out loud. The Coen's have made a magnificiant film again. But I think
the first time you watch this movie, you get to know the characters. The second
time, now that you know them, you laugh sooo hard it could hurt you. I strongly
would reccomend ANYONE seeing this because if you are not, you are truly missing
a film gem for the ages. 10/10
5. This movie spends most of its time preaching that it is the script that makes
the movie, but apparently there was no script when they shot this waste of time!
The trailer makes this out to be a comedy, but the film can't decide if it wants
to be a comedy, a drama, a romance or an action film. Press releases indicated
that Shatner and Hamlin made this movie because they loved the script (what were
they thinking?). If you like William Shatner (I do) see "Free Enterprise"
instead.
```

[34]:

```python
# Define a function to compute metrics
def compute_metrics(preds, labels):
    preds = torch.argmax(torch.tensor(preds), axis=1)
    accuracy = accuracy_score(labels, preds)
    precision, recall, f1, _ = precision_recall_fscore_support(labels, preds,␣
 ↪average='binary')
    return {"accuracy": accuracy, "precision": precision, "recall": recall,␣
 ↪"f1": f1}

# (b) ROC Curve and AUC metric
def plot_roc_curve(y_true, y_scores):
    fpr, tpr, _ = roc_curve(y_true, y_scores)
    roc_auc = auc(fpr, tpr)
    plt.figure()
    plt.plot(fpr, tpr, color='darkorange', lw=2, label='ROC curve (area = %0.
 ↪2f)' % roc_auc)
    plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
    plt.xlim([0.0, 1.0])
    plt.ylim([0.0, 1.05])
    plt.xlabel('False Positive Rate')
    plt.ylabel('True Positive Rate')
    plt.title('Receiver Operating Characteristic')
    plt.legend(loc="lower right")
    plt.show()
    return roc_auc
```

[62]:

```
# Complete code for Fine-tuning BERT for sentiment classification
# Load the tokenizer and model
tokenizer = DistilBertTokenizer.from_pretrained('distilbert-base-uncased')
model = DistilBertForSequenceClassification.
 ↪from_pretrained('distilbert-base-uncased', num_labels=2).to(device)

# Tokenize the dataset
def tokenize_function(examples):
    return tokenizer(examples['text'], truncation=True, padding=True,␣
 ↪max_length=512)

train_data = train_dataset.map(tokenize_function, batched=True)
test_data = test_dataset.map(tokenize_function, batched=True)

# Set the format for PyTorch
train_data.set_format(type='torch', columns=['input_ids', 'attention_mask',␣
 ↪'label'])
test_data.set_format(type='torch', columns=['input_ids', 'attention_mask',␣
 ↪'label'])
```

```
/opt/conda/lib/python3.10/site-
packages/transformers/tokenization_utils_base.py:1601: FutureWarning:
`clean_up_tokenization_spaces` was not set. It will be set to `True` by default.
This behavior will be depracted in transformers v4.45, and will be then set to
`False` by default. For more details check this issue:
https://github.com/huggingface/transformers/issues/31884
  warnings.warn(
Some weights of DistilBertForSequenceClassification were not initialized from
the model checkpoint at distilbert-base-uncased and are newly initialized:
['classifier.bias', 'classifier.weight', 'pre_classifier.bias',
'pre_classifier.weight']
You should probably TRAIN this model on a down-stream task to be able to use it
for predictions and inference.
```

3.1 (a) Fine-tune with frozen layers (except the last layer)

[63]:

```
training_args = TrainingArguments(
    output_dir='./results',
    evaluation_strategy="epoch",
    per_device_train_batch_size=8,
    per_device_eval_batch_size=8,
    num_train_epochs=3,
    weight_decay=0.01,
    logging_dir='./logs',
)
```

```python
trainer = Trainer(
    model=model,
    args=training_args,
    train_dataset=train_data,
    eval_dataset=test_data,
    compute_metrics=lambda p: compute_metrics(p.predictions, p.label_ids),
)
```

```
/opt/conda/lib/python3.10/site-packages/transformers/training_args.py:1525:
FutureWarning: `evaluation_strategy` is deprecated and will be removed in
version 4.46 of  Transformers. Use `eval_strategy` instead
  warnings.warn(
```

[64]:
```python
for param in model.distilbert.parameters():
    param.requires_grad = False  # Freeze all layers except the last␣
 ↪classification layer

# Ensure classifier parameters are trainable
for param in model.classifier.parameters():
    param.requires_grad = True  # Keep the classification layer trainable

# Train the model
trainer.train()
```

```
<IPython.core.display.HTML object>
```

[64]:
```
TrainOutput(global_step=375, training_loss=0.6365109456380208,
metrics={'train_runtime': 56.876, 'train_samples_per_second': 52.746,
'train_steps_per_second': 6.593, 'total_flos': 397402195968000.0, 'train_loss':
0.6365109456380208, 'epoch': 3.0})
```

[65]:
```python
# Evaluate and output metrics
eval_results = trainer.evaluate()
print(f"Evaluation Results (All Layers): {eval_results}")
```

```
<IPython.core.display.HTML object>
```

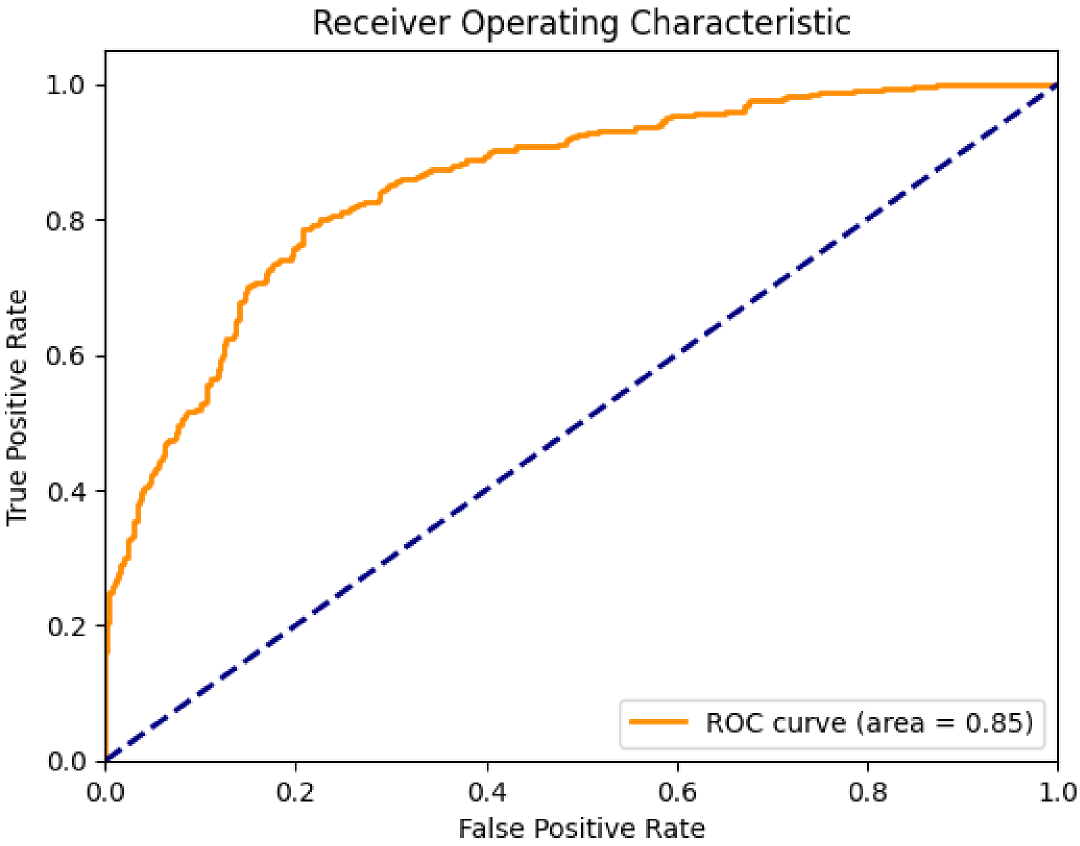
```
Evaluation Results (All Layers): {'eval_loss': 0.6042457222938538,
'eval_accuracy': 0.779, 'eval_precision': 0.779874213836478, 'eval_recall':
0.7622950819672131, 'eval_f1': 0.7709844559585491, 'eval_runtime': 8.8255,
'eval_samples_per_second': 113.309, 'eval_steps_per_second': 14.164, 'epoch':
3.0}
```

3.2 (b) Metrics for the fine-tuned model

[66]:

```
import matplotlib.pyplot as plt

print(f"Evaluation Accuracy: {eval_results['eval_accuracy']}")
print(f"Evaluation Loss: {eval_results['eval_loss']}")
print(f"Evaluation Precision: {eval_results['eval_precision']}")
print(f"Evaluation Recall: {eval_results['eval_recall']}")
print(f"Evaluation F1 Score: {eval_results['eval_f1']}")

# Get model predictions
predictions = trainer.predict(test_data).predictions
predicted_probabilities = torch.softmax(torch.tensor(predictions), dim=1)[:, 1]
true_labels = test_data['label']

# Plot ROC curve and compute AUC
auc_value = plot_roc_curve(true_labels, predicted_probabilities)
print(f"AUC: {auc_value}")
```

```
Evaluation Accuracy: 0.779
Evaluation Loss: 0.6042457222938538
Evaluation Precision: 0.779874213836478
Evaluation Recall: 0.7622950819672131
Evaluation F1 Score: 0.7709844559585491
```



```
AUC: 0.8536236872438524
```

3.3 (c) Fine-tune all layers and compare results

```
[52]: model = DistilBertForSequenceClassification.
       ↪from_pretrained('distilbert-base-uncased', num_labels=2).to(device)
```

```
Some weights of DistilBertForSequenceClassification were not initialized from
the model checkpoint at distilbert-base-uncased and are newly initialized:
['classifier.bias', 'classifier.weight', 'pre_classifier.bias',
'pre_classifier.weight']
You should probably TRAIN this model on a down-stream task to be able to use it
for predictions and inference.
```

```
[56]: # Unfreeze all layers
      for param in model.distilbert.parameters():
          param.requires_grad = True

      # Re-train with all layers fine-tuned
      trainer.train()
```

```
<IPython.core.display.HTML object>
```

[56]:

```
TrainOutput(global_step=375, training_loss=0.5274981689453125,
metrics={'train_runtime': 108.0455, 'train_samples_per_second': 27.766,
'train_steps_per_second': 3.471, 'total_flos': 397402195968000.0, 'train_loss':
0.5274981689453125, 'epoch': 3.0})
```

[57]:

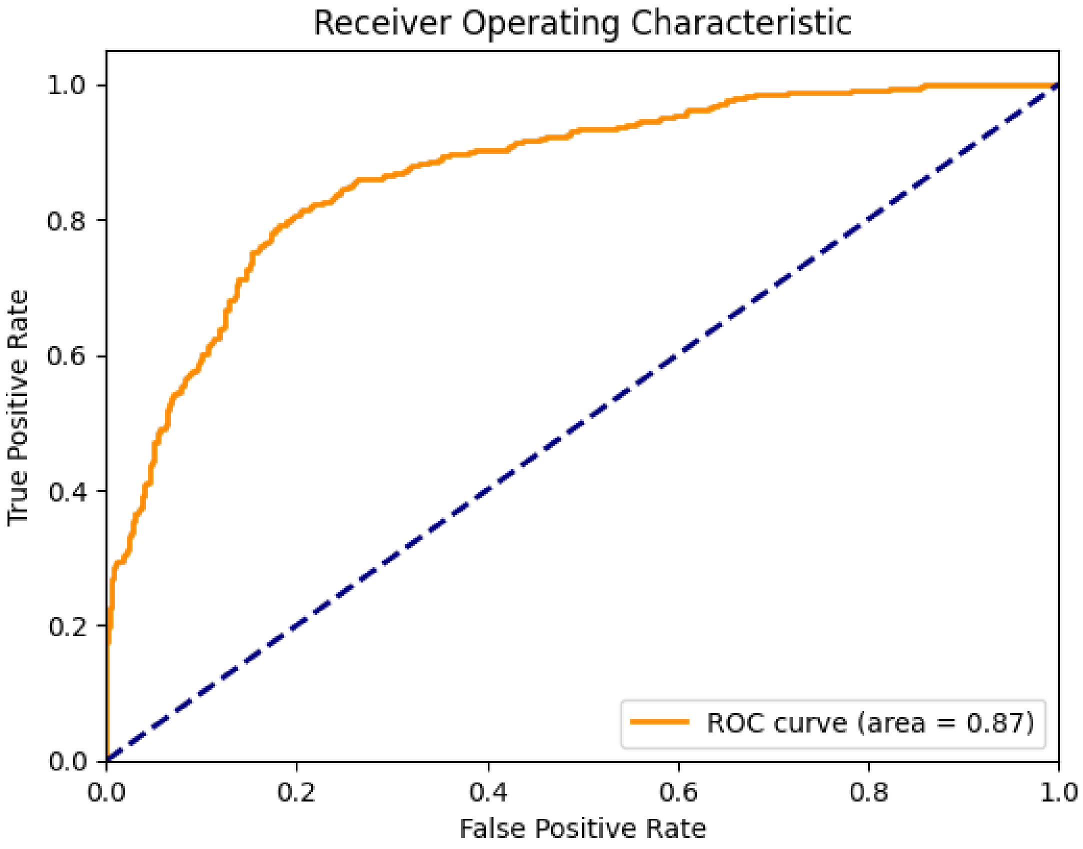
```python
# Evaluate and output metrics again
eval_results_all_layers = trainer.evaluate()
print(f"Evaluation Results (All Layers): {eval_results_all_layers}")
```

```
<IPython.core.display.HTML object>

Evaluation Results (All Layers): {'eval_loss': 0.5214744806289673,
'eval_accuracy': 0.803, 'eval_precision': 0.8089171974522293, 'eval_recall':
0.7807377049180327, 'eval_f1': 0.794577685088634, 'eval_runtime': 8.8032,
'eval_samples_per_second': 113.595, 'eval_steps_per_second': 14.199, 'epoch':
3.0}
```

[58]:

```python
print(f"Evaluation Accuracy: {eval_results_all_layers['eval_accuracy']}")
print(f"Evaluation Loss: {eval_results_all_layers['eval_loss']}")
print(f"Evaluation Precision: {eval_results_all_layers['eval_precision']}")
print(f"Evaluation Recall: {eval_results_all_layers['eval_recall']}")
print(f"Evaluation F1 Score: {eval_results_all_layers['eval_f1']}")

# Get model predictions
predictions = trainer.predict(test_data).predictions
predicted_probabilities = torch.softmax(torch.tensor(predictions), dim=1)[:, 1]
true_labels = test_data['label']

# Plot ROC curve and compute AUC
auc_value = plot_roc_curve(true_labels, predicted_probabilities)
print(f"AUC: {auc_value}")
```

```
Evaluation Accuracy: 0.803
Evaluation Loss: 0.5214744806289673
Evaluation Precision: 0.8089171974522293
Evaluation Recall: 0.7807377049180327
Evaluation F1 Score: 0.794577685088634
```



```
AUC: 0.8692887102971312
```

Training all layers results in a lower training loss and evaluation loss than that of training only the last layer of the model. This performance is expected as it allowed the model to adapt to this specific task. Further, the model has more parameters to optimize, increasing model capacity compared to the case when we only optimize parameters from the last layer.