recruit3

June 6, 2024

DP

[4]:

```
!pip install ja-ginza
```

```
Collecting ja-ginza
  Using cached ja_ginza-5.2.0-py3-none-any.whl.metadata (5.8 kB)
Collecting spacy<4.0.0,>=3.4.4 (from ja-ginza)
  Using cached
spacy-3.7.5-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata
(27 kB)
Collecting sudachipy<0.7.0,>=0.6.2 (from ja-ginza)
  Using cached SudachiPy-0.6.8-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (12 kB)
Collecting sudachidict-core>=20210802 (from ja-ginza)
  Using cached SudachiDict_core-20240409-py3-none-any.whl.metadata (2.5 kB)
Collecting ginza<5.3.0,>=5.2.0 (from ja-ginza)
  Using cached ginza-5.2.0-py3-none-any.whl.metadata (448 bytes)
Collecting plac>=1.3.3 (from ginza<5.3.0,>=5.2.0->ja-ginza)
  Using cached plac-1.4.3-py2.py3-none-any.whl.metadata (5.9 kB)
Collecting spacy-legacy<3.1.0,>=3.0.11 (from spacy<4.0.0,>=3.4.4->ja-ginza)
  Using cached spacy_legacy-3.0.12-py2.py3-none-any.whl.metadata (2.8 kB)
Collecting spacy-loggers<2.0.0,>=1.0.0 (from spacy<4.0.0,>=3.4.4->ja-ginza)
  Using cached spacy_loggers-1.0.5-py3-none-any.whl.metadata (23 kB)
Collecting murmurhash<1.1.0,>=0.28.0 (from spacy<4.0.0,>=3.4.4->ja-ginza)
  Using cached murmurhash-1.0.10-cp311-cp311-
manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux2014_x86_6
4.whl.metadata (2.0 kB)
Collecting cymem<2.1.0,>=2.0.2 (from spacy<4.0.0,>=3.4.4->ja-ginza)
  Using cached
cymem-2.0.8-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata
(8.4 kB)
Collecting preshed<3.1.0,>=3.0.2 (from spacy<4.0.0,>=3.4.4->ja-ginza)
  Using cached preshed-3.0.9-cp311-cp311-
manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux2014_x86_6
4.whl.metadata (2.2 kB)
Collecting thinc<8.3.0,>=8.2.2 (from spacy<4.0.0,>=3.4.4->ja-ginza)
  Using cached
thinc-8.2.4-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata
(15 kB)
```

```
Collecting wasabi<1.2.0,>=0.9.1 (from spacy<4.0.0,>=3.4.4->ja-ginza)
  Using cached wasabi-1.1.3-py3-none-any.whl.metadata (28 kB)
Collecting srsly<3.0.0,>=2.4.3 (from spacy<4.0.0,>=3.4.4->ja-ginza)
  Using cached
srsly-2.4.8-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata
(20 kB)
Collecting catalogue<2.1.0,>=2.0.6 (from spacy<4.0.0,>=3.4.4->ja-ginza)
  Using cached catalogue-2.0.10-py3-none-any.whl.metadata (14 kB)
Collecting weasel<0.5.0,>=0.1.0 (from spacy<4.0.0,>=3.4.4->ja-ginza)
  Using cached weasel-0.4.1-py3-none-any.whl.metadata (4.6 kB)
Requirement already satisfied: typer<1.0.0,>=0.3.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (0.12.3)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (4.66.2)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (2.31.0)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (2.7.0)
Requirement already satisfied: jinja2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (3.1.3)
Requirement already satisfied: setuptools in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (69.5.1)
Requirement already satisfied: packaging>=20.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (23.2)
Collecting langcodes<4.0.0,>=3.2.0 (from spacy<4.0.0,>=3.4.4->ja-ginza)
  Using cached langcodes-3.4.0-py3-none-any.whl.metadata (29 kB)
Requirement already satisfied: numpy>=1.19.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (1.26.4)
Collecting language-data>=1.2 (from
langcodes<4.0.0,>=3.2.0->spacy<4.0.0,>=3.4.4->ja-ginza)
  Using cached language_data-1.2.0-py3-none-any.whl.metadata (4.3 kB)
Requirement already satisfied: annotated-types>=0.4.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4->spacy<4.0.0,>=3.4.4->ja-ginza) (0.6.0)
Requirement already satisfied: pydantic-core==2.18.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4->spacy<4.0.0,>=3.4.4->ja-ginza) (2.18.1)
Requirement already satisfied: typing-extensions>=4.6.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4->spacy<4.0.0,>=3.4.4->ja-ginza) (4.11.0)
```

```
Requirement already satisfied: charset-normalizer<4,>=2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
requests<3.0.0,>=2.13.0->spacy<4.0.0,>=3.4.4->ja-ginza) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
requests<3.0.0,>=2.13.0->spacy<4.0.0,>=3.4.4->ja-ginza) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
requests<3.0.0,>=2.13.0->spacy<4.0.0,>=3.4.4->ja-ginza) (1.26.18)
Requirement already satisfied: certifi>=2017.4.17 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
requests<3.0.0,>=2.13.0->spacy<4.0.0,>=3.4.4->ja-ginza) (2024.2.2)
Collecting blis<0.8.0,>=0.7.8 (from
thinc<8.3.0,>=8.2.2->spacy<4.0.0,>=3.4.4->ja-ginza)
  Using cached
blis-0.7.11-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata
(7.4 kB)
Collecting confection<1.0.0,>=0.0.1 (from
thinc<8.3.0,>=8.2.2->spacy<4.0.0,>=3.4.4->ja-ginza)
  Using cached confection-0.1.5-py3-none-any.whl.metadata (19 kB)
Requirement already satisfied: click>=8.0.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
typer<1.0.0,>=0.3.0->spacy<4.0.0,>=3.4.4->ja-ginza) (8.1.7)
Requirement already satisfied: shellingham>=1.3.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
typer<1.0.0,>=0.3.0->spacy<4.0.0,>=3.4.4->ja-ginza) (1.5.4)
Requirement already satisfied: rich>=10.11.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
typer<1.0.0,>=0.3.0->spacy<4.0.0,>=3.4.4->ja-ginza) (13.7.1)
Collecting cloudpathlib<1.0.0,>=0.7.0 (from
weasel<0.5.0,>=0.1.0->spacy<4.0.0,>=3.4.4->ja-ginza)
  Using cached cloudpathlib-0.18.1-py3-none-any.whl.metadata (14 kB)
Collecting smart-open<8.0.0,>=5.2.1 (from
weasel<0.5.0,>=0.1.0->spacy<4.0.0,>=3.4.4->ja-ginza)
  Downloading smart_open-7.0.4-py3-none-any.whl.metadata (23 kB)
Requirement already satisfied: MarkupSafe>=2.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
jinja2->spacy<4.0.0,>=3.4.4->ja-ginza) (2.1.5)
Collecting marisa-trie>=0.7.7 (from language-
data>=1.2->langcodes<4.0.0,>=3.2.0->spacy<4.0.0,>=3.4.4->ja-ginza)
  Downloading marisa_trie-1.2.0-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (8.7 kB)
Requirement already satisfied: markdown-it-py>=2.2.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy<4.0.0,>=3.4.4->ja-ginza) (3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy<4.0.0,>=3.4.4->ja-ginza) (2.17.2)
```

```
Requirement already satisfied: wrapt in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from smart-
open<8.0.0,>=5.2.1->weasel<0.5.0,>=0.1.0->spacy<4.0.0,>=3.4.4->ja-ginza)
(1.16.0)
Requirement already satisfied: mdurl~=0.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from markdown-it-
py>=2.2.0->rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy<4.0.0,>=3.4.4->ja-ginza)
(0.1.2)
Using cached ja_ginza-5.2.0-py3-none-any.whl (59.1 MB)
Using cached ginza-5.2.0-py3-none-any.whl (21 kB)
Using cached
spacy-3.7.5-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (6.6 MB)
Using cached SudachiDict_core-20240409-py3-none-any.whl (72.0 MB)
Using cached
SudachiPy-0.6.8-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (2.6
MB)
Using cached catalogue-2.0.10-py3-none-any.whl (17 kB)
Using cached
cymem-2.0.8-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (46 kB)
Using cached langcodes-3.4.0-py3-none-any.whl (182 kB)
Using cached murmurhash-1.0.10-cp311-cp311-
manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux2014_x86_6
4.whl (29 kB)
Using cached plac-1.4.3-py2.py3-none-any.whl (22 kB)
Using cached preshed-3.0.9-cp311-cp311-
manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux2014_x86_6
4.whl (157 kB)
Using cached spacy_legacy-3.0.12-py2.py3-none-any.whl (29 kB)
Using cached spacy_loggers-1.0.5-py3-none-any.whl (22 kB)
Using cached
srsly-2.4.8-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (490 kB)
Using cached
thinc-8.2.4-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (920 kB)
Using cached wasabi-1.1.3-py3-none-any.whl (27 kB)
Using cached weasel-0.4.1-py3-none-any.whl (50 kB)
Using cached
blis-0.7.11-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (10.2 MB)
Using cached cloudpathlib-0.18.1-py3-none-any.whl (47 kB)
Using cached confection-0.1.5-py3-none-any.whl (35 kB)
Using cached language_data-1.2.0-py3-none-any.whl (5.4 MB)
Downloading smart_open-7.0.4-py3-none-any.whl (61 kB)

61.2/61.2 kB 4.1 MB/s eta 0:00:00
Downloading
marisa_trie-1.2.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl
(1.4 MB)

1.4/1.4 MB 16.6 MB/s eta 0:00:0031m18.3 MB/s eta
```

```
0:00:01
Installing collected packages: sudachipy, plac, cymem, wasabi,
sudachidict-core, spacy-loggers, spacy-legacy, smart-open, murmurhash, marisa-
trie, cloudpathlib, catalogue, blis, srsly, preshed, language-data, langcodes,
confection, weasel, thinc, spacy, ginza, ja-ginza
Successfully installed blis-0.7.11 catalogue-2.0.10 cloudpathlib-0.18.1
confection-0.1.5 cymem-2.0.8 ginza-5.2.0 ja-ginza-5.2.0 langcodes-3.4.0
language-data-1.2.0 marisa-trie-1.2.0 murmurhash-1.0.10 plac-1.4.3 preshed-3.0.9
smart-open-7.0.4 spacy-3.7.5 spacy-legacy-3.0.12 spacy-loggers-1.0.5 srsly-2.4.8
sudachidict-core-20240409 sudachipy-0.6.8 thinc-8.2.4 wasabi-1.1.3 weasel-0.4.1
```

0.1 1.

[5]:

```
!pip install spacy
!pip install sklearn
!pip install python-dp
```

```
Requirement already satisfied: spacy in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (3.7.5)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(1.0.10)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(2.0.8)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(3.0.9)
Requirement already satisfied: thinc<8.3.0,>=8.2.2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(8.2.4)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(1.1.3)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(2.4.8)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(2.0.10)
Requirement already satisfied: weasel<0.5.0,>=0.1.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
```

```
(0.4.1)
Requirement already satisfied: typer<1.0.0,>=0.3.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(0.12.3)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(4.66.2)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(2.31.0)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(2.7.0)
Requirement already satisfied: jinja2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(3.1.3)
Requirement already satisfied: setuptools in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(69.5.1)
Requirement already satisfied: packaging>=20.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(23.2)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(3.4.0)
Requirement already satisfied: numpy>=1.19.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(1.26.4)
Requirement already satisfied: language-data>=1.2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
langcodes<4.0.0,>=3.2.0->spacy) (1.2.0)
Requirement already satisfied: annotated-types>=0.4.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4->spacy) (0.6.0)
Requirement already satisfied: pydantic-core==2.18.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4->spacy) (2.18.1)
Requirement already satisfied: typing-extensions>=4.6.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4->spacy) (4.11.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
requests<3.0.0,>=2.13.0->spacy) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
requests<3.0.0,>=2.13.0->spacy) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
requests<3.0.0,>=2.13.0->spacy) (1.26.18)
Requirement already satisfied: certifi>=2017.4.17 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
requests<3.0.0,>=2.13.0->spacy) (2024.2.2)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
thinc<8.3.0,>=8.2.2->spacy) (0.7.11)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
thinc<8.3.0,>=8.2.2->spacy) (0.1.5)
Requirement already satisfied: click>=8.0.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
typer<1.0.0,>=0.3.0->spacy) (8.1.7)
Requirement already satisfied: shellingham>=1.3.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
typer<1.0.0,>=0.3.0->spacy) (1.5.4)
Requirement already satisfied: rich>=10.11.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
typer<1.0.0,>=0.3.0->spacy) (13.7.1)
Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
weasel<0.5.0,>=0.1.0->spacy) (0.18.1)
Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
weasel<0.5.0,>=0.1.0->spacy) (7.0.4)
Requirement already satisfied: MarkupSafe>=2.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
jinja2->spacy) (2.1.5)
Requirement already satisfied: marisa-trie>=0.7.7 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from language-
data>=1.2->langcodes<4.0.0,>=3.2.0->spacy) (1.2.0)
Requirement already satisfied: markdown-it-py>=2.2.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (2.17.2)
Requirement already satisfied: wrapt in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from smart-
open<8.0.0,>=5.2.1->weasel<0.5.0,>=0.1.0->spacy) (1.16.0)
Requirement already satisfied: mdurl~=0.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from markdown-it-
py>=2.2.0->rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (0.1.2)
Collecting sklearn
  Using cached sklearn-0.0.post12.tar.gz (2.6 kB)
  Installing build dependencies … done
  Getting requirements to build wheel … error
  error: subprocess-exited-with-error
```

```
 × Getting requirements to build wheel did not run
successfully.
  exit code: 1
  > [15 lines of output]
   The 'sklearn' PyPI package is deprecated, use 'scikit-learn'
   rather than 'sklearn' for pip commands.

   Here is how to fix this error in the main use cases:
   - use 'pip install scikit-learn' rather than 'pip install
sklearn'
   - replace 'sklearn' by 'scikit-learn' in your pip requirements
files
    (requirements.txt, setup.py, setup.cfg, Pipfile, etc …)
   - if the 'sklearn' package is used by one of your dependencies,
    it would be great if you take some time to track which package
uses
    'sklearn' instead of 'scikit-learn' and report it to their
issue tracker
   - as a last resort, set the environment variable
    SKLEARN_ALLOW_DEPRECATED_SKLEARN_PACKAGE_INSTALL=True to avoid
this error

   More information is available at
   https://github.com/scikit-learn/sklearn-pypi-package
   [end of output]

 note: This error originates from a subprocess, and is likely not a
problem with pip.
error: subprocess-exited-with-error

× Getting requirements to build wheel did not run
successfully.
 exit code: 1
 > See above for output.

note: This error originates from a subprocess, and is likely not a
problem with pip.
Collecting python-dp
 Using cached python_dp-1.1.4-cp311-cp311-manylinux1_x86_64.whl.metadata (5.1
kB)
Using cached python_dp-1.1.4-cp311-cp311-manylinux1_x86_64.whl (3.8 MB)
Installing collected packages: python-dp
Successfully installed python-dp-1.1.4
```

0.2 2.

[6]:

```
import pandas as pd

data_path = './data/reviews_with_sentiment.csv'

df = pd.read_csv(data_path)
df
```

[6]:

```
                                               review sentiment
0                                                 neutral
1                                                   neutral
2                                                 positive
3                                              positive
4                                                positive
…                                               …       …
5548                              negative
5549                                     negative
5550                              neutral
5551                             negative
5552                                      …   neutral

[5553 rows x 2 columns]
```

[7]:

```
from sklearn.feature_extraction.text import CountVectorizer
import spacy
import itertools
from typing import List, Tuple
# spaCy
nlp = spacy.load('ja_ginza')

#
POS = ['ADJ', 'ADV', 'INTJ', 'PROPN', 'NOUN', 'VERB']
MAX_TERMS_IN_DOC = 5
NGRAM = 1
MAX_DF = 1.0
MIN_DF = 0.0
NUM_VOCAB = 10000
TOP_K = 20

def flatten(*lists) -> list:
    res = []
    for l in list(itertools.chain.from_iterable(lists)):
        for e in l:
            res.append(e)
    return res

def remove_duplicates(l: List[Tuple[str, float]]) -> List[Tuple[str, float]]:
    d = {}
    for e in l:
        d[e[0]] = e[1]
    return list(d.items())

#
df["doc"] = [nlp(review) for review in df["review"]]
```

```
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-
packages/torch/cuda/__init__.py:118: UserWarning: CUDA initialization: CUDA
unknown error - this may be due to an incorrectly set up environment, e.g.
changing env variable CUDA_VISIBLE_DEVICES after program start. Setting the
available devices to be zero. (Triggered internally at
../c10/cuda/CUDAFunctions.cpp:108.)
  return torch._C._cuda_getDeviceCount() > 0
```

0.3 3. Bag-of-Words

[8]:
```
bows = {}
cvs = {}

for sentiment in df["sentiment"].unique():
    tokens = []
    for doc in df[df["sentiment"] == sentiment]["doc"]:
        similarities = [(token.similarity(doc), token.lemma_) for token in doc␣
  ↪if token.pos_ in POS]
        similarities = remove_duplicates(similarities)
        similarities = sorted(similarities, key=lambda sim: sim[1],␣
  ↪reverse=True)[:MAX_TERMS_IN_DOC]
        tokens.append([similarity[1] for similarity in similarities])

    cv = CountVectorizer(ngram_range=(1, NGRAM), max_df=MAX_DF, min_df=MIN_DF,␣
  ↪max_features=NUM_VOCAB)
    bows[sentiment] = cv.fit_transform(flatten(tokens)).toarray()
    cvs[sentiment] = cv
```

```
/tmp/ipykernel_255084/4107160084.py:7: UserWarning: [W008] Evaluating
Token.similarity based on empty vectors.
  similarities = [(token.similarity(doc), token.lemma_) for token in doc if
token.pos_ in POS]
```

[ ]:
```
!pip install numpy
```

0.4 4.

```python
from pydp.algorithms.laplacian import Count
import numpy as np#
vocabs = {}
term_frequencies = {}

for sentiment in df["sentiment"].unique():
    bow = bows[sentiment]
    cv = cvs[sentiment]

    vocab = cv.vocabulary_
    term_frequency = np.sum(bow, axis=0)
    vocabs[sentiment] = vocab
    term_frequencies[sentiment] = term_frequency

    indices_topk = np.argsort(term_frequency)[::-1][:TOP_K]
    bow_topk = np.take(bow, indices_topk, axis=1)
    reverse_vocab = {vocab[k]: k for k in vocab.keys()}
    words = [reverse_vocab[i] for i in indices_topk]

    print(sentiment, ":")
    for w, c in zip(words, term_frequency[indices_topk]):
        print(w, ":", c)
```

```
neutral :
  : 216
  : 93
  : 67
  : 60
  : 59
  : 50
  : 47
  : 45
  : 43
  : 43
   : 34
   : 32
  : 30
  : 27
  : 27
  : 25
  : 25
  : 23
  : 22
  : 21
positive :
  : 463
  : 315
   : 206
  : 198
  : 193
  : 181
  : 175
  : 144
  : 141
  : 140
   : 124
  : 106
  : 102
  : 99
  : 96
  : 90
  : 90
  : 87
  : 84
   : 80
negative :
  : 99
  : 97
  : 48
  : 42
  : 39
  : 36
  : 34
  : 33
  : 27
  : 24
  : 22
  : 19
   : 16
  : 16
   : 16
  : 16
  : 15
   : 15
  : 15
  : 14
```

0.5 5.

[12]:

```
from pydp.algorithms.laplacian import Count

def preprocess_for_private_counts(tf: np.ndarray) -> List[np.ndarray]:
    repeated_words = []
    for i, term in enumerate(tf):
        repeated_words.append(np.repeat(i, term))
    return repeated_words

def cal_private_count(epsilon: float, max_partition_contributed: float,␣
 ↪max_contributions_per_partition: float, repeated_words: List[np.ndarray]) ->␣
 ↪List[int]:
    private_counts = []
    for repeated_word in repeated_words:
        counter = Count(epsilon, max_partition_contributed,␣
 ↪max_contributions_per_partition)
        count = counter.quick_result(repeated_word)
        private_counts.append(count)
    return private_counts

def top_k_words_and_counts(k: int, tf: np.ndarray, vocab: dict) ->␣
 ↪List[Tuple[str, int]]:
    indices_topk = np.argsort(tf)[::-1][:k]
    reverse_vocab = {vocab[key]: key for key in vocab.keys()}
    words = [reverse_vocab[i] for i in indices_topk]
    counts = [tf[i] for i in indices_topk]
    return list(zip(words, counts))

epsilons = [0.01, 0.05, 0.1, 0.3, 0.7, 1.0, 2.0, 3.0, 7.0, 10.0]
MAX_DUPLICATED_TERMS = 1

#
results = {sentiment: {"no DP": None, **{eps: None for eps in epsilons}} for␣
 ↪sentiment in df["sentiment"].unique()}

for eps in epsilons:
    print(" : ", eps)
    for sentiment in df["sentiment"].unique():
        repeated_words =␣
 ↪preprocess_for_private_counts(term_frequencies[sentiment])
        private_counts = cal_private_count(eps, MAX_TERMS_IN_DOC,␣
 ↪MAX_DUPLICATED_TERMS, repeated_words)
        words_and_counts = top_k_words_and_counts(TOP_K, private_counts,␣
 ↪vocabs[sentiment])
        results[sentiment][eps] = words_and_counts
        print(sentiment, ":")
        print(words_and_counts)

#
for sentiment in df["sentiment"].unique():
    words_and_counts = top_k_words_and_counts(TOP_K,␣
↪term_frequencies[sentiment], vocabs[sentiment])
    results[sentiment]["no DP"] = words_and_counts
```

```
 :  0.01
neutral :
[(' ', 2694), (' ', 2607), (' ', 2385), (' ', 2352), ('  ', 2264), ('  ',
1918), (' ', 1902), ('  ', 1876), (' ', 1858), ('  ', 1843), ('   ',
1827), (' ', 1806), ('  ', 1761), (' ', 1738), (' ', 1682), ('  ', 1674),
('  ', 1663), (' ', 1633), (' ', 1611), (' ', 1607)]
positive :
[('  ', 2957), ('  ', 2954), (' ', 2922), ('  ', 2532), (' ', 2465),
('  ', 2432), (' ', 2332), (' ', 2267), (' ', 2206), (' ', 2115),
('   ', 1992), ('  ', 1989), (' ', 1917), (' ', 1860), ('   ', 1851),
('  ', 1848), (' ', 1811), (' ', 1802), ('  ', 1798), (' ', 1798)]
negative :
[('  ', 3678), (' ', 3487), (' ', 2567), (' ', 2447), ('  ', 2402), (' ',
2327), (' ', 2210), ('  ', 2087), ('  ', 2074), (' ', 1980), ('  ',
1980), (' ', 1888), ('  ', 1851), (' ', 1832), ('  ', 1832), (' ', 1827),
(' ', 1817), ('  ', 1811), ('  ', 1738), ('  ', 1682)]
 :  0.05
neutral :
[(' ', 715), (' ', 697), ('  ', 685), ('  ', 680), (' ', 507), (' ',
460), (' ', 457), (' ', 455), (' ', 451), (' ', 440), ('  ', 427), ('  ',
425), (' ', 424), ('  ', 420), (' ', 419), (' ', 406), ('  ', 400),
(' ', 383), ('  ', 383), ('  ', 383)]
positive :
[(' ', 637), (' ', 616), (' ', 614), ('  ', 585), (' ', 562), (' ', 554),
(' ', 534), (' ', 526), (' ', 486), (' ', 472), (' ', 465), (' ', 460),
(' ', 458), (' ', 445), (' ', 444), ('  ', 428), ('  ', 406), ('   ',
387), ('  ', 375), ('  ', 366)]
negative :
[(' ', 766), (' ', 553), (' ', 405), (' ', 365), ('  ', 358), (' ', 337),
('  ', 330), ('  ', 320), ('  ', 313), (' ', 309), (' ', 295), (' ',
288), (' ', 288), (' ', 281), ('  ', 277), (' ', 267), (' ', 266), ('  ',
264), (' ', 259), ('  ', 255)]
 :  0.1
neutral :
[(' ', 357), (' ', 298), ('    ', 286), (' ', 241), ('  ', 240), (' ',
236), (' ', 233), (' ', 221), (' ', 209), (' ', 204), (' ', 197), (' ',
195), ('  ', 193), ('  ', 190), (' ', 190), (' ', 190), ('  ', 189),
('     ', 183), (' ', 178), ('  ', 172)]
positive :
[(' ', 431), ('  ', 301), (' ', 298), (' ', 284), (' ', 272), (' ', 269),
('  ', 251), (' ', 245), ('  ', 243), (' ', 241), (' ', 232), (' ',
227), (' ', 226), (' ', 221), (' ', 218), (' ', 216), ('  ', 215), (' ',
211), (' ', 211), (' ', 210)]
```

```
negative :
[(' ', 371), ('  ', 331), ('  ', 314), (' ', 255), (' ', 252), (' ', 245),
(' ', 217), ('  ', 203), (' ', 197), (' ', 196), ('  ', 188), (' ', 171),
('   ', 169), ('  ', 166), (' ', 163), ('  ', 161), (' ', 160), ('  ',
158), ('    ', 156), (' ', 153)]
 :  0.3
neutral :
[(' ', 262), ('  ', 123), (' ', 107), (' ', 107), (' ', 99), ('  ', 77),
(' ', 73), (' ', 72), (' ', 67), (' ', 66), (' ', 65), (' ', 64), (' ',
63), ('  ', 63), (' ', 63), (' ', 62), (' ', 62), (' ', 60), (' ', 59),
('  ', 59)]
positive :
[(' ', 499), (' ', 316), ('  ', 213), (' ', 209), (' ', 197), ('  ',
156), (' ', 155), (' ', 155), (' ', 152), (' ', 140), (' ', 117), (' ',
113), ('  ', 107), (' ', 107), (' ', 106), ('  ', 103), (' ', 95),
('  ', 94), (' ', 92), (' ', 89)]
negative :
[(' ', 121), (' ', 101), ('  ', 94), (' ', 93), (' ', 93), (' ', 78),
(' ', 76), ('  ', 72), (' ', 72), (' ', 71), ('  ', 71), (' ', 70),
('   ', 65), (' ', 64), ('   ', 63), (' ', 61), ('  ', 61), ('  ',
59), ('  ', 58), ('  ', 57)]
 :  0.7
neutral :
[(' ', 250), (' ', 105), (' ', 74), (' ', 68), (' ', 67), (' ', 56),
(' ', 56), (' ', 52), (' ', 51), (' ', 48), (' ', 47), ('  ', 42), (' ',
41), ('  ', 40), (' ', 37), ('  ', 36), (' ', 36), ('   ', 35), (' ',
34), ('  ', 34)]
positive :
[(' ', 474), (' ', 311), ('  ', 210), (' ', 183), (' ', 182), (' ', 179),
(' ', 165), (' ', 148), (' ', 142), (' ', 139), ('  ', 136), (' ', 112),
(' ', 105), (' ', 105), ('  ', 97), (' ', 94), (' ', 93), ('  ', 89),
(' ', 89), (' ', 89)]
negative :
[(' ', 109), (' ', 100), (' ', 53), (' ', 49), (' ', 42), (' ', 41),
(' ', 40), (' ', 39), ('  ', 38), (' ', 37), (' ', 34), (' ', 34), (' ',
34), (' ', 31), ('  ', 31), (' ', 31), (' ', 30), (' ', 29), ('  ', 28),
(' ', 28)]
 :  1.0
neutral :
[(' ', 215), (' ', 94), (' ', 61), (' ', 57), (' ', 56), (' ', 55), (' ',
52), (' ', 43), (' ', 43), (' ', 42), (' ', 42), ('  ', 35), ('  ', 34),
(' ', 30), (' ', 30), (' ', 28), (' ', 28), (' ', 27), ('  ', 26), (' ',
26)]
positive :
[(' ', 456), (' ', 318), ('  ', 210), (' ', 202), (' ', 198), (' ', 187),
(' ', 167), (' ', 148), (' ', 138), (' ', 133), (' ', 124), ('  ', 120),
(' ', 107), (' ', 103), (' ', 100), (' ', 94), ('  ', 91), (' ', 91),
(' ', 89), ('  ', 87)]
```

```
negative :
[(' ', 104), (' ', 97), (' ', 54), (' ', 40), (' ', 39), (' ', 34), (' ',
34), (' ', 32), (' ', 29), (' ', 29), ('  ', 28), (' ', 28), ('  ', 27),
('   ', 26), (' ', 26), ('  ', 25), ('  ', 25), ('  ', 25), (' ', 24),
(' ', 24)]
 :  2.0
neutral :
[(' ', 215), (' ', 93), (' ', 67), (' ', 60), (' ', 60), (' ', 49), (' ',
46), (' ', 46), (' ', 46), (' ', 44), ('  ', 34), (' ', 33), (' ', 32),
('  ', 31), (' ', 30), (' ', 27), ('  ', 25), (' ', 25), (' ', 24), (' ',
24)]
positive :
[(' ', 463), (' ', 315), ('  ', 205), (' ', 196), (' ', 187), (' ', 179),
(' ', 176), (' ', 145), (' ', 137), (' ', 131), ('  ', 124), (' ', 108),
(' ', 100), (' ', 98), (' ', 94), (' ', 93), (' ', 90), (' ', 90), (' ',
84), (' ', 84)]
negative :
[(' ', 99), (' ', 97), (' ', 48), (' ', 42), (' ', 41), (' ', 36), (' ',
34), (' ', 32), (' ', 28), (' ', 25), (' ', 24), (' ', 22), (' ', 20),
('  ', 19), ('  ', 17), ('  ', 17), (' ', 17), ('  ', 16), ('   ', 16),
(' ', 15)]
 :  3.0
neutral :
[(' ', 217), (' ', 99), (' ', 67), (' ', 59), (' ', 59), (' ', 48), (' ',
46), (' ', 46), (' ', 41), (' ', 40), ('  ', 35), ('  ', 35), (' ', 32),
(' ', 28), (' ', 27), (' ', 26), (' ', 25), (' ', 24), (' ', 22), (' ',
22)]
positive :
[(' ', 463), (' ', 315), ('  ', 207), (' ', 200), (' ', 195), (' ', 183),
(' ', 178), (' ', 151), (' ', 143), (' ', 139), ('  ', 123), (' ', 106),
(' ', 101), (' ', 101), (' ', 96), (' ', 88), (' ', 88), (' ', 87), (' ',
85), (' ', 83)]
negative :
[(' ', 99), (' ', 94), (' ', 48), (' ', 40), (' ', 34), (' ', 33), (' ',
32), (' ', 30), (' ', 27), (' ', 22), (' ', 21), (' ', 21), ('  ', 18),
(' ', 16), (' ', 16), (' ', 16), ('  ', 16), (' ', 16), ('  ', 15), (' ',
15)]
 :  7.0
neutral :
[(' ', 215), (' ', 92), (' ', 66), (' ', 60), (' ', 59), (' ', 49), (' ',
46), (' ', 45), (' ', 43), (' ', 43), ('  ', 34), ('  ', 32), (' ', 31),
(' ', 29), (' ', 25), (' ', 25), (' ', 24), (' ', 23), (' ', 22), (' ',
21)]
positive :
[(' ', 463), (' ', 314), ('  ', 203), (' ', 197), (' ', 193), (' ', 181),
(' ', 175), (' ', 147), (' ', 141), (' ', 139), ('  ', 124), (' ', 106),
(' ', 101), (' ', 99), (' ', 96), (' ', 90), (' ', 90), (' ', 87), (' ',
84), ('  ', 81)]
```

```
negative :
[('  ', 99), ('  ', 97), ('  ', 49), ('  ', 43), ('  ', 38), ('  ', 35), ('  ',
33), ('  ', 33), ('  ', 26), ('  ', 25), ('  ', 21), ('  ', 19), ('  ', 17),
('   ', 17), ('  ', 16), ('  ', 16), ('    ', 16), ('  ', 16), ('  ', 15),
('  ', 15)]
 :  10.0
neutral :
[('  ', 217), ('  ', 93), ('  ', 67), ('  ', 59), ('  ', 59), ('  ', 52), ('  ',
47), ('  ', 43), ('  ', 43), ('  ', 43), ('   ', 34), ('   ', 32), ('  ', 30),
('  ', 27), ('  ', 26), ('  ', 25), ('  ', 25), ('  ', 23), ('  ', 22), ('  ',
22)]
positive :
[('  ', 463), ('  ', 314), ('    ', 206), ('  ', 198), ('  ', 193), ('  ', 181),
('  ', 174), ('  ', 143), ('  ', 140), ('  ', 140), ('   ', 124), ('  ', 106),
('  ', 102), ('  ', 98), ('  ', 98), ('  ', 90), ('  ', 90), ('  ', 87), ('  ',
83), ('   ', 79)]
negative :
[('  ', 100), ('  ', 97), ('  ', 48), ('  ', 43), ('  ', 39), ('  ', 36), ('  ',
34), ('  ', 34), ('  ', 28), ('  ', 24), ('  ', 22), ('  ', 19), ('  ', 18),
('   ', 18), ('  ', 16), ('    ', 16), ('  ', 15), ('  ', 15), ('   ', 14),
('  ', 14)]
```

0.6 6.

[14]:
```python
def display_results(results):
    for sentiment, data in results.items():
        print(f"\nSentiment: {sentiment}\n")
        df = pd.DataFrame(columns=["rank", "word", "count"] + [f" ={eps}" for␣
↪eps in epsilons])
        for rank, (word, count) in enumerate(data["no DP"], start=1):
            row = {"rank": rank, "word": word, "count": count}
            for eps in epsilons:
                if eps in data and rank <= len(data[eps]):
                    row[f" ={eps}"] = data[eps][rank-1][1] if rank-1 <␣
↪len(data[eps]) else None
            df = pd.concat([df, pd.DataFrame([row])], ignore_index=True)
        print(df)

display_results(results)
```

```
Sentiment: neutral

   rank word count  =0.01  =0.05  =0.1  =0.3  =0.7  =1.0  =2.0  =3.0  =7.0  \
0     1        216   2694    715   357   262   250   215   215   217   215
1     2         93   2607    697   298   123   105    94    93    99    92
2     3         67   2385    685   286   107    74    61    67    67    66
3     4         60   2352    680   241   107    68    57    60    59    60
```

```
4      5          59   2264    507   240    99    67    56    60    59    59
5      6          50   1918    460   236    77    56    55    49    48    49
6      7          47   1902    457   233    73    56    52    46    46    46
7      8          45   1876    455   221    72    52    43    46    46    45
8      9          43   1858    451   209    67    51    43    46    41    43
9     10          43   1843    440   204    66    48    42    44    40    43
10    11         34   1827    427   197    65    47    42    34    35    34
11    12         32   1806    425   195    64    42    35    33    35    32
12    13          30   1761    424   193    63    41    34    32    32    31
13    14          27   1738    420   190    63    40    30    31    28    29
14    15          27   1682    419   190    63    37    30    30    27    25
15    16          25   1674    406   190    62    36    28    27    26    25
16    17          25   1663    400   189    62    36    28    25    25    24
17    18          23   1633    383   183    60    35    27    25    24    23
18    19          22   1611    383   178    59    34    26    24    22    22
19    20          21   1607    383   172    59    34    26    24    22    21

   =10.0
0     217
1      93
2      67
3      59
4      59
5      52
6      47
7      43
8      43
9      43
10     34
11     32
12     30
13     27
14     26
15     25
16     25
17     23
18     22
19     22

Sentiment: positive

   rank  word count  =0.01  =0.05  =0.1  =0.3  =0.7  =1.0  =2.0  =3.0  =7.0  \
0     1          463   2957    637   431   499   474   456   463   463   463
1     2          315   2954    616   301   316   311   318   315   315   314
2     3         206   2922    614   298   213   210   210   205   207   203
3     4          198   2532    585   284   209   183   202   196   200   197
4     5          193   2465    562   272   197   182   198   187   195   193
5     6          181   2432    554   269   156   179   187   179   183   181
```

```
6      7          175   2332    534   251   155   165   167   176   178   175
7      8          144   2267    526   245   155   148   148   145   151   147
8      9          141   2206    486   243   152   142   138   137   143   141
9     10          140   2115    472   241   140   139   133   131   139   139
10    11          124   1992    465   232   117   136   124   124   123   124
11    12          106   1989    460   227   113   112   120   108   106   106
12    13          102   1917    458   226   107   105   107   100   101   101
13    14           99   1860    445   221   107   105   103    98   101    99
14    15           96   1851    444   218   106    97   100    94    96    96
15    16           90   1848    428   216   103    94    94    93    88    90
16    17           90   1811    406   215    95    93    91    90    88    90
17    18           87   1802    387   211    94    89    91    90    87    87
18    19           84   1798    375   211    92    89    89    84    85    84
19    20           80   1798    366   210    89    89    87    84    83    81

   =10.0
0     463
1     314
2     206
3     198
4     193
5     181
6     174
7     143
8     140
9     140
10    124
11    106
12    102
13     98
14     98
15     90
16     90
17     87
18     83
19     79

Sentiment: negative

   rank  word count  =0.01  =0.05  =0.1  =0.3  =0.7  =1.0  =2.0  =3.0  =7.0  \
0     1          99   3678    766   371   121   109   104    99    99    99
1     2          97   3487    553   331   101   100    97    97    94    97
2     3          48   2567    405   314    94    53    54    48    48    49
3     4          42   2447    365   255    93    49    40    42    40    43
4     5          39   2402    358   252    93    42    39    41    34    38
5     6          36   2327    337   245    78    41    34    36    33    35
6     7          34   2210    330   217    76    40    34    34    32    33
7     8          33   2087    320   203    72    39    32    32    30    33
```

```
8     9         27   2074    313    197     72     38     29     28     27     26
9    10         24   1980    309    196     71     37     29     25     22     25
10   11         22   1980    295    188     71     34     28     24     21     21
11   12         19   1888    288    171     70     34     28     22     21     19
12   13         16   1851    288    169     65     34     27     20     18     17
13   14         16   1832    281    166     64     31     26     19     16     17
14   15        16   1832    277    163     63     31     26     17     16     16
15   16         16   1827    267    161     61     31     25     17     16     16
16   17         15   1817    266    160     61     30     25     17     16     16
17   18         15   1811    264    158     59     29     25     16     16     16
18   19         15   1738    259    156     58     28     24     16     15     15
19   20         14   1682    255    153     57     28     24     15     15     15

    =10.0
0      100
1       97
2       48
3       43
4       39
5       36
6       34
7       34
8       28
9       24
10      22
11      19
12      18
13      18
14      16
15      16
16      15
17      15
18      14
19      14
```

0.7 7.

[20]:
```
!pip install matplotlib
!pip install seaborn
```

```
Requirement already satisfied: matplotlib in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (3.8.4)
Requirement already satisfied: contourpy>=1.0.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from matplotlib)
(1.2.1)
Requirement already satisfied: cycler>=0.10 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from matplotlib)
(0.12.1)
```

```
Requirement already satisfied: fonttools>=4.22.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from matplotlib)
(4.51.0)
Requirement already satisfied: kiwisolver>=1.3.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from matplotlib)
(1.4.5)
Requirement already satisfied: numpy>=1.21 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from matplotlib)
(1.26.4)
Requirement already satisfied: packaging>=20.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from matplotlib)
(23.2)
Requirement already satisfied: pillow>=8 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from matplotlib)
(10.3.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from matplotlib)
(3.1.2)
Requirement already satisfied: python-dateutil>=2.7 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from matplotlib)
(2.9.0.post0)
Requirement already satisfied: six>=1.5 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from python-
dateutil>=2.7->matplotlib) (1.16.0)
Collecting seaborn
  Downloading seaborn-0.13.2-py3-none-any.whl.metadata (5.4 kB)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from seaborn)
(1.26.4)
Requirement already satisfied: pandas>=1.2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from seaborn)
(2.2.2)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from seaborn)
(3.8.4)
Requirement already satisfied: contourpy>=1.0.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (1.2.1)
Requirement already satisfied: cycler>=0.10 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (4.51.0)
Requirement already satisfied: kiwisolver>=1.3.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (1.4.5)
Requirement already satisfied: packaging>=20.0 in
```

```
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (23.2)
Requirement already satisfied: pillow>=8 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (10.3.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (3.1.2)
Requirement already satisfied: python-dateutil>=2.7 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
pandas>=1.2->seaborn) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
pandas>=1.2->seaborn) (2024.1)
Requirement already satisfied: six>=1.5 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from python-
dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.16.0)
Downloading seaborn-0.13.2-py3-none-any.whl (294 kB)

294.9/294.9 kB 5.3 MB/s eta 0:00:00[31m4.6 MB/s
eta 0:00:01
Installing collected packages: seaborn
Successfully installed seaborn-0.13.2
```

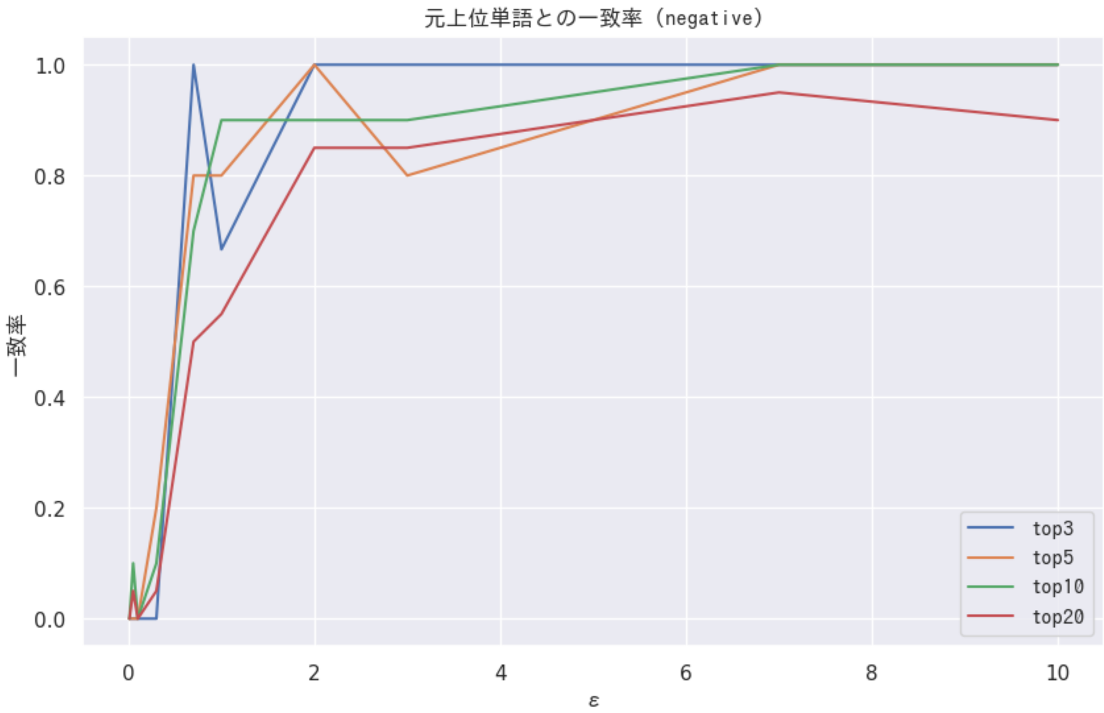
[29]:

```python
import matplotlib.pyplot as plt
import matplotlib.font_manager as fm
import numpy as np

#
font_path = '/usr/share/fonts/truetype/fonts-japanese-gothic.ttf'  #
font_prop = fm.FontProperties(fname=font_path)

def calculate_match_rate(original_top_k, dp_top_k):
    match_count = len(set(original_top_k) & set(dp_top_k))
    return match_count / len(original_top_k) if original_top_k else 0

def plot_match_rate(results, epsilons, sentiment_label):
    top_k_values = [3, 5, 10, 20]
    match_rates = {k: [] for k in top_k_values}

    for eps in epsilons:
        for k in top_k_values:
            original_top_k = [word for word, count in␣
 ↪results[sentiment_label]["no DP"][:k]]
```

```
            dp_top_k = [word for word, count in results[sentiment_label][eps][:
↪k]]
            match_rate = calculate_match_rate(original_top_k, dp_top_k)
            match_rates[k].append(match_rate)

    plt.figure(figsize=(10, 6))
    for k, rates in match_rates.items():
        plt.plot(epsilons, rates, label=f'top{k}')

    plt.xlabel(' ', fontproperties=font_prop)
    plt.ylabel('  ', fontproperties=font_prop)
    plt.title(f'      ({sentiment_label})', fontproperties=font_prop)
    plt.legend(prop=font_prop)
    plt.grid(True)
    plt.show()

epsilons = [0.01, 0.05, 0.1, 0.3, 0.7, 1.0, 2.0, 3.0, 7.0, 10.0]

plot_match_rate(results, epsilons, 'negative')
plot_match_rate(results, epsilons, 'neutral')
plot_match_rate(results, epsilons, 'positive')
```

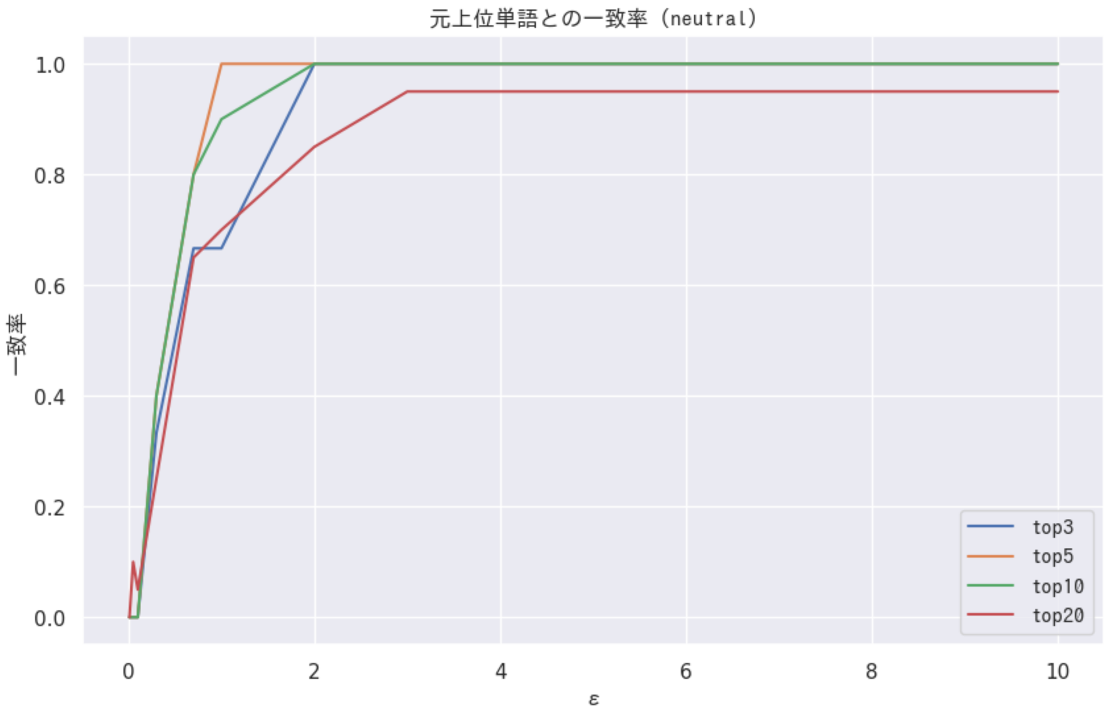
```
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
```

```
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
```

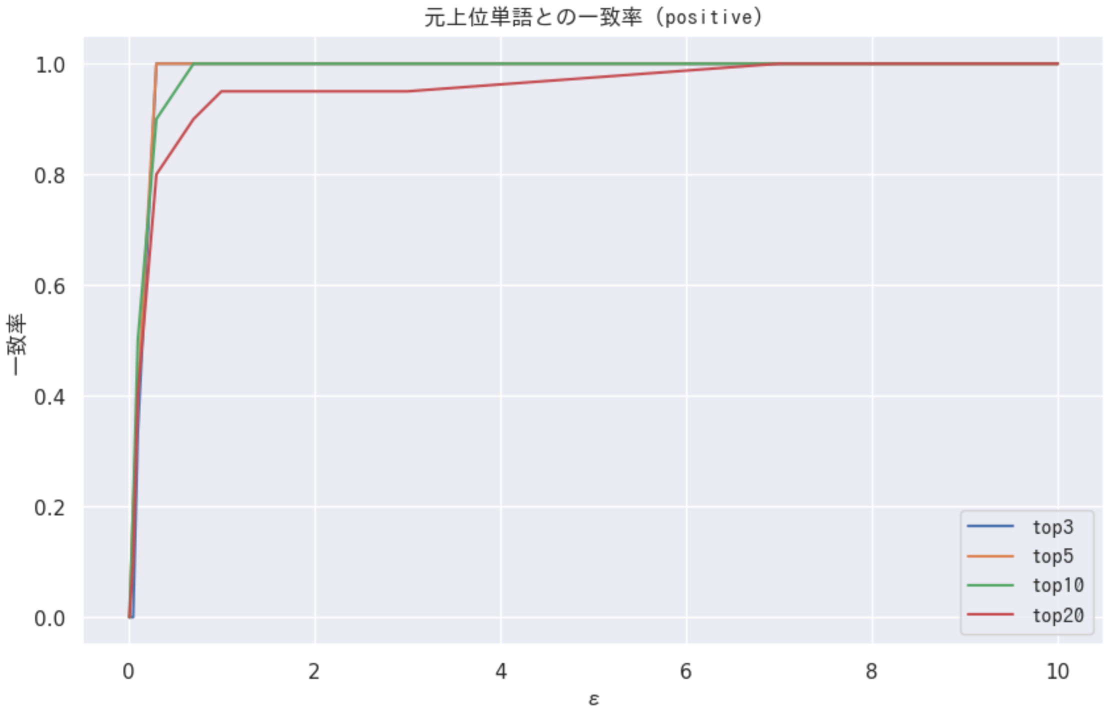


```
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
```

```
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
```



```
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
```

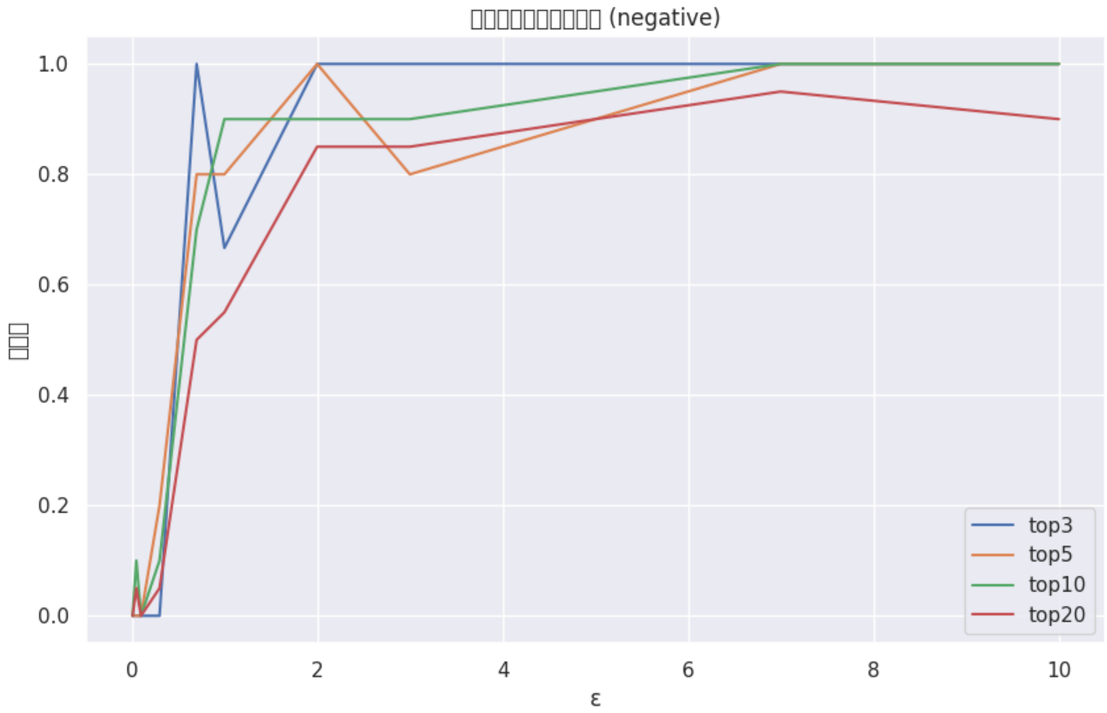


[25]:
```
import matplotlib.pyplot as plt
import matplotlib.font_manager as fm
import numpy as np

#
plt.rcParams['font.family'] = 'IPAGothic'

def calculate_match_rate(original_top_k, dp_top_k):
    match_count = len(set(original_top_k) & set(dp_top_k))
    return match_count / len(original_top_k) if original_top_k else 0

def plot_match_rate(results, epsilons, sentiment_label):
    top_k_values = [3, 5, 10, 20]
    match_rates = {k: [] for k in top_k_values}

    for eps in epsilons:
        for k in top_k_values:
            original_top_k = [word for word, count in␣
 ↪results[sentiment_label]["no DP"][:k]]
            dp_top_k = [word for word, count in results[sentiment_label][eps][:
 ↪k]]
            match_rate = calculate_match_rate(original_top_k, dp_top_k)
            match_rates[k].append(match_rate)
    plt.figure(figsize=(10, 6))
    for k, rates in match_rates.items():
        plt.plot(epsilons, rates, label=f'top{k}')

    plt.xlabel(' ')
    plt.ylabel('  ')
    plt.title(f'      ({sentiment_label})')
    plt.legend()
    plt.grid(True)
    plt.show()

epsilons = [0.01, 0.05, 0.1, 0.3, 0.7, 1.0, 2.0, 3.0, 7.0, 10.0]

plot_match_rate(results, epsilons, 'negative')
plot_match_rate(results, epsilons, 'neutral')
plot_match_rate(results, epsilons, 'positive')
```

```
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
```

```
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
```

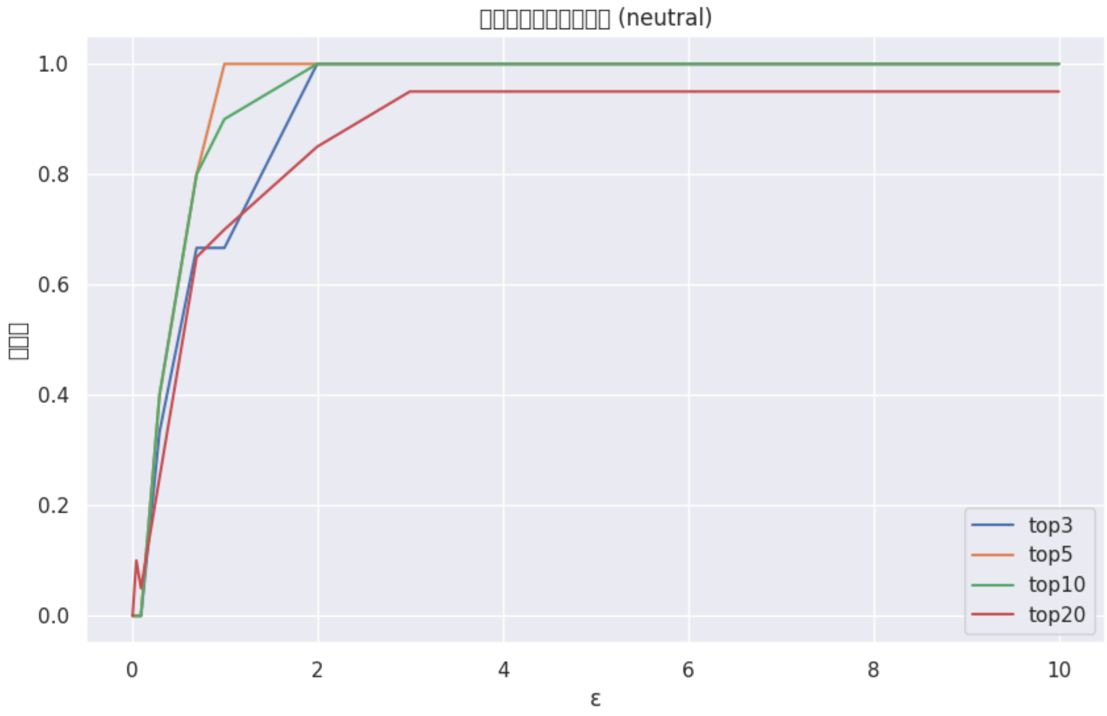
```
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
```



```
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
```

```
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
```

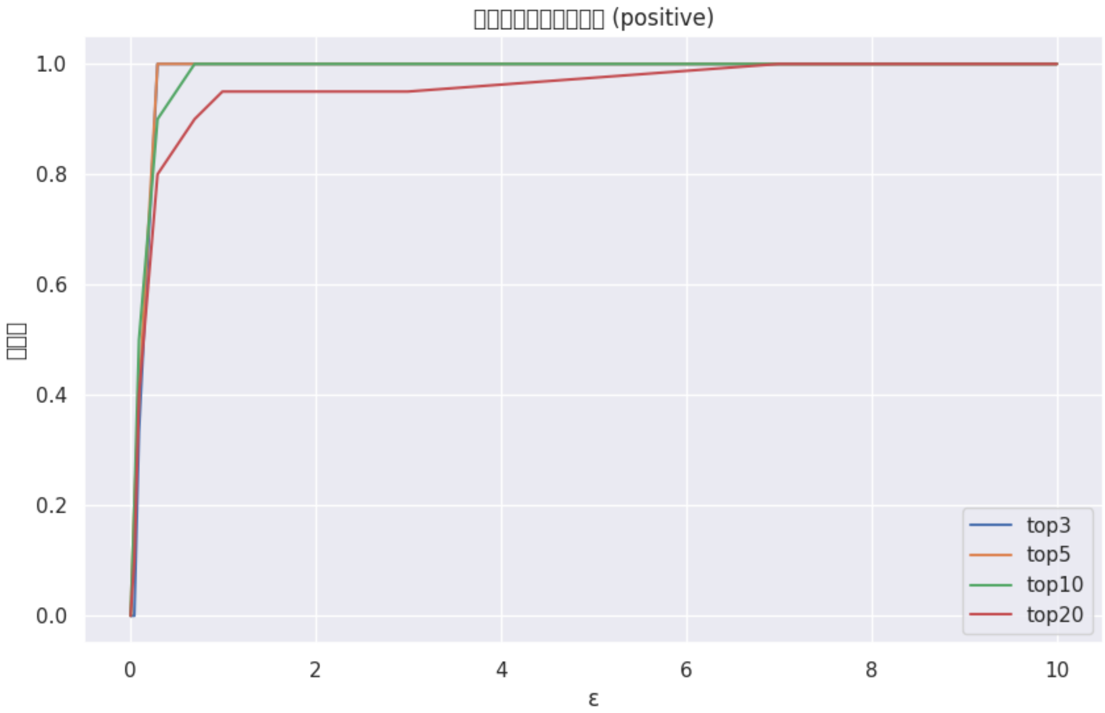
```
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
```



```
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
```

```
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
```

```
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
findfont: Font family 'IPAGothic' not found.
```



0.8 : ML

[26]:

```
!pip install diffprivlib
```

```
Collecting diffprivlib
  Downloading diffprivlib-0.6.4-py3-none-any.whl.metadata (9.6 kB)
Requirement already satisfied: numpy>=1.21.6 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from diffprivlib)
(1.26.4)
Requirement already satisfied: scikit-learn>=0.24.2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from diffprivlib)
(1.4.2)
Requirement already satisfied: scipy>=1.7.3 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from diffprivlib)
(1.13.0)
Requirement already satisfied: joblib>=0.16.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from diffprivlib)
(1.4.2)
Requirement already satisfied: setuptools>=49.0.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from diffprivlib)
(69.5.1)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from scikit-
learn>=0.24.2->diffprivlib) (3.5.0)
```

```
Downloading diffprivlib-0.6.4-py3-none-any.whl (176 kB)

176.5/176.5 kB 3.4 MB/s eta 0:00:00[36m0:00:01
Installing collected packages: diffprivlib
Successfully installed diffprivlib-0.6.4
```

[ ]:

```python
import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB as SklearnGaussianNB
import numpy as np
import spacy
from typing import List, Tuple
import itertools
import matplotlib.pyplot as plt
from diffprivlib.models import GaussianNB as DPGaussianNB

#
data_path = 'path/to/reviews_with_sentiment.csv'  #
df = pd.read_csv(data_path)

# spaCy
nlp = spacy.load('ja_ginza')

#
POS = ['ADJ', 'ADV', 'INTJ', 'PROPN', 'NOUN', 'VERB']
MAX_TERMS_IN_DOC = 5
NGRAM = 1
MAX_DF = 1.0
MIN_DF = 0.01
NUM_VOCAB = 10000

def flatten(*lists) -> list:
    res = []
    for l in list(itertools.chain.from_iterable(lists)):
        for e in l:
            res.append(e)
    return res

def remove_duplicates(l: List[Tuple[str, float]]) -> List[Tuple[str, float]]:
    d = {}
    for e in l:
        d[e[0]] = e[1]
    return list(d.items())

#    BoW
tokens = []
```

```python
for doc in df["review"]:
    parsed_doc = nlp(doc)
    similarities = [(token.similarity(parsed_doc), token.lemma_) for token in parsed_doc if token.pos_ in POS]
    similarities = remove_duplicates(similarities)
    similarities = sorted(similarities, key=lambda sim: sim[1], reverse=True)[:MAX_TERMS_IN_DOC]
    tokens.append([similarity[1] for similarity in similarities])

cv = CountVectorizer(ngram_range=(1, NGRAM), max_df=MAX_DF, min_df=MIN_DF, max_features=NUM_VOCAB)
bow = cv.fit_transform([" ".join(ts) for ts in tokens]).toarray()

#
m = {
    "positive": 1,
    "neutral": 0,
    "negative": 0,
}
df["sentiment"] = df["sentiment"].map(m)
df["bow"] = bow.tolist()

X_train, X_test, y_train, y_test = train_test_split(df["bow"], df["sentiment"], test_size=0.2)
X_train = [list(x) for x in X_train]
X_test = [list(x) for x in X_test]

#
clf = SklearnGaussianNB()
clf.fit(X_train, y_train)
print("Non-DP accuracy: ", clf.score(X_test, y_test))

#
epsilons = np.logspace(-2, 2, 50)
dim = np.array(X_train).shape[1]
lowers = np.zeros(dim)
uppers = np.ones(dim)
accuracies = {}

for epsilon in epsilons:
    accuracy = []
    for _ in range(20):
        dp_clf = DPGaussianNB(bounds=(lowers, uppers), epsilon=epsilon)
        dp_clf.fit(X_train, y_train)
        accuracy.append(dp_clf.score(X_test, y_test))
    accuracies[epsilon] = accuracy
```

```
#
x = epsilons
y = [np.mean(accuracies[eps]) for eps in epsilons]
e = [np.std(accuracies[eps]) for eps in epsilons]

plt.figure(figsize=(10, 6))
plt.semilogx(x, y)
plt.errorbar(x, y, yerr=e, marker='o', capthick=1, capsize=10, lw=1)
plt.xlabel(' ')
plt.ylabel('accuracy')
plt.ylim(0, 1)
plt.title('       accuracy  ')
plt.grid(True)
plt.show()
```