Assignment_2_A

May 1, 2025

1 ΕΠΕΞΕΡΓΑΣΙΑ ΦΥΣΙΚΗΣ ΓΛΩΣΣΑΣ - Ε 2

1.1 A. Word embeddings

Μ : Ε Φ Γ

Σ : Ι Κ

Ε : 2025-04-12 | v.0.0.1

1.1.1 ΠΡΟΕΤΟΙΜΑΣΙΑ ΠΕΡΙΒΑΛΛΟΝΤΟΣ

1.2 Ε Β

Σ : - `gensim`: word embeddings - `scikit-learn`: t-SNE - `plotly`: - `numpy`, `pandas`:

[1]:
```
%pip install gensim scikit-learn plotly pandas notebook ipywidgets nbformat␣
↪--upgrade
```

```
Requirement already satisfied: gensim in ./.conda/lib/python3.11/site-packages
(4.3.3)
Requirement already satisfied: scikit-learn in ./.conda/lib/python3.11/site-
packages (1.6.1)
Requirement already satisfied: plotly in ./.conda/lib/python3.11/site-packages
(6.0.1)
Requirement already satisfied: pandas in ./.conda/lib/python3.11/site-packages
(2.2.3)
Requirement already satisfied: notebook in ./.conda/lib/python3.11/site-packages
(7.4.0)
Requirement already satisfied: ipywidgets in ./.conda/lib/python3.11/site-
packages (8.1.6)
Requirement already satisfied: nbformat in ./.conda/lib/python3.11/site-packages
(5.10.4)
Requirement already satisfied: numpy<2.0,>=1.18.5 in
./.conda/lib/python3.11/site-packages (from gensim) (1.26.4)
Requirement already satisfied: scipy<1.14.0,>=1.7.0 in
./.conda/lib/python3.11/site-packages (from gensim) (1.13.1)
Requirement already satisfied: smart-open>=1.8.1 in
./.conda/lib/python3.11/site-packages (from gensim) (7.1.0)
```

```
Requirement already satisfied: joblib>=1.2.0 in ./.conda/lib/python3.11/site-
packages (from scikit-learn) (1.4.2)
Requirement already satisfied: threadpoolctl>=3.1.0 in
./.conda/lib/python3.11/site-packages (from scikit-learn) (3.6.0)
Requirement already satisfied: narwhals>=1.15.1 in ./.conda/lib/python3.11/site-
packages (from plotly) (1.34.1)
Requirement already satisfied: packaging in ./.conda/lib/python3.11/site-
packages (from plotly) (24.2)
Requirement already satisfied: python-dateutil>=2.8.2 in
./.conda/lib/python3.11/site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in ./.conda/lib/python3.11/site-
packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in ./.conda/lib/python3.11/site-
packages (from pandas) (2025.2)
Requirement already satisfied: jupyter-server<3,>=2.4.0 in
./.conda/lib/python3.11/site-packages (from notebook) (2.15.0)
Requirement already satisfied: jupyterlab-server<3,>=2.27.1 in
./.conda/lib/python3.11/site-packages (from notebook) (2.27.3)
Requirement already satisfied: jupyterlab<4.5,>=4.4.0rc0 in
./.conda/lib/python3.11/site-packages (from notebook) (4.4.0)
Requirement already satisfied: notebook-shim<0.3,>=0.2 in
./.conda/lib/python3.11/site-packages (from notebook) (0.2.4)
Requirement already satisfied: tornado>=6.2.0 in ./.conda/lib/python3.11/site-
packages (from notebook) (6.4.2)
Requirement already satisfied: comm>=0.1.3 in ./.conda/lib/python3.11/site-
packages (from ipywidgets) (0.2.1)
Requirement already satisfied: ipython>=6.1.0 in ./.conda/lib/python3.11/site-
packages (from ipywidgets) (8.30.0)
Requirement already satisfied: traitlets>=4.3.1 in ./.conda/lib/python3.11/site-
packages (from ipywidgets) (5.14.3)
Requirement already satisfied: widgetsnbextension~=4.0.14 in
./.conda/lib/python3.11/site-packages (from ipywidgets) (4.0.14)
Requirement already satisfied: jupyterlab_widgets~=3.0.14 in
./.conda/lib/python3.11/site-packages (from ipywidgets) (3.0.14)
Requirement already satisfied: fastjsonschema>=2.15 in
./.conda/lib/python3.11/site-packages (from nbformat) (2.21.1)
Requirement already satisfied: jsonschema>=2.6 in ./.conda/lib/python3.11/site-
packages (from nbformat) (4.23.0)
Requirement already satisfied: jupyter-core!=5.0.*,>=4.12 in
./.conda/lib/python3.11/site-packages (from nbformat) (5.7.2)
Requirement already satisfied: decorator in ./.conda/lib/python3.11/site-
packages (from ipython>=6.1.0->ipywidgets) (5.1.1)
Requirement already satisfied: jedi>=0.16 in ./.conda/lib/python3.11/site-
packages (from ipython>=6.1.0->ipywidgets) (0.19.2)
Requirement already satisfied: matplotlib-inline in
./.conda/lib/python3.11/site-packages (from ipython>=6.1.0->ipywidgets) (0.1.6)
Requirement already satisfied: prompt-toolkit<3.1.0,>=3.0.41 in
./.conda/lib/python3.11/site-packages (from ipython>=6.1.0->ipywidgets) (3.0.43)
```

```
Requirement already satisfied: pygments>=2.4.0 in ./.conda/lib/python3.11/site-
packages (from ipython>=6.1.0->ipywidgets) (2.15.1)
Requirement already satisfied: stack-data in ./.conda/lib/python3.11/site-
packages (from ipython>=6.1.0->ipywidgets) (0.2.0)
Requirement already satisfied: typing-extensions>=4.6 in
./.conda/lib/python3.11/site-packages (from ipython>=6.1.0->ipywidgets) (4.12.2)
Requirement already satisfied: pexpect>4.3 in ./.conda/lib/python3.11/site-
packages (from ipython>=6.1.0->ipywidgets) (4.8.0)
Requirement already satisfied: attrs>=22.2.0 in ./.conda/lib/python3.11/site-
packages (from jsonschema>=2.6->nbformat) (25.3.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in
./.conda/lib/python3.11/site-packages (from jsonschema>=2.6->nbformat)
(2024.10.1)
Requirement already satisfied: referencing>=0.28.4 in
./.conda/lib/python3.11/site-packages (from jsonschema>=2.6->nbformat) (0.36.2)
Requirement already satisfied: rpds-py>=0.7.1 in ./.conda/lib/python3.11/site-
packages (from jsonschema>=2.6->nbformat) (0.24.0)
Requirement already satisfied: platformdirs>=2.5 in
./.conda/lib/python3.11/site-packages (from jupyter-
core!=5.0.*,>=4.12->nbformat) (4.3.7)
Requirement already satisfied: anyio>=3.1.0 in ./.conda/lib/python3.11/site-
packages (from jupyter-server<3,>=2.4.0->notebook) (4.9.0)
Requirement already satisfied: argon2-cffi>=21.1 in
./.conda/lib/python3.11/site-packages (from jupyter-server<3,>=2.4.0->notebook)
(23.1.0)
Requirement already satisfied: jinja2>=3.0.3 in ./.conda/lib/python3.11/site-
packages (from jupyter-server<3,>=2.4.0->notebook) (3.1.6)
Requirement already satisfied: jupyter-client>=7.4.4 in
./.conda/lib/python3.11/site-packages (from jupyter-server<3,>=2.4.0->notebook)
(8.6.3)
Requirement already satisfied: jupyter-events>=0.11.0 in
./.conda/lib/python3.11/site-packages (from jupyter-server<3,>=2.4.0->notebook)
(0.12.0)
Requirement already satisfied: jupyter-server-terminals>=0.4.4 in
./.conda/lib/python3.11/site-packages (from jupyter-server<3,>=2.4.0->notebook)
(0.5.3)
Requirement already satisfied: nbconvert>=6.4.4 in ./.conda/lib/python3.11/site-
packages (from jupyter-server<3,>=2.4.0->notebook) (7.16.6)
Requirement already satisfied: overrides>=5.0 in ./.conda/lib/python3.11/site-
packages (from jupyter-server<3,>=2.4.0->notebook) (7.7.0)
Requirement already satisfied: prometheus-client>=0.9 in
./.conda/lib/python3.11/site-packages (from jupyter-server<3,>=2.4.0->notebook)
(0.21.1)
Requirement already satisfied: pyzmq>=24 in ./.conda/lib/python3.11/site-
packages (from jupyter-server<3,>=2.4.0->notebook) (26.2.0)
Requirement already satisfied: send2trash>=1.8.2 in
./.conda/lib/python3.11/site-packages (from jupyter-server<3,>=2.4.0->notebook)
(1.8.3)
```

```
Requirement already satisfied: terminado>=0.8.3 in ./.conda/lib/python3.11/site-
packages (from jupyter-server<3,>=2.4.0->notebook) (0.18.1)
Requirement already satisfied: websocket-client>=1.7 in
./.conda/lib/python3.11/site-packages (from jupyter-server<3,>=2.4.0->notebook)
(1.8.0)
Requirement already satisfied: async-lru>=1.0.0 in ./.conda/lib/python3.11/site-
packages (from jupyterlab<4.5,>=4.4.0rc0->notebook) (2.0.5)
Requirement already satisfied: httpx>=0.25.0 in ./.conda/lib/python3.11/site-
packages (from jupyterlab<4.5,>=4.4.0rc0->notebook) (0.28.1)
Requirement already satisfied: ipykernel>=6.5.0 in ./.conda/lib/python3.11/site-
packages (from jupyterlab<4.5,>=4.4.0rc0->notebook) (6.29.5)
Requirement already satisfied: jupyter-lsp>=2.0.0 in
./.conda/lib/python3.11/site-packages (from jupyterlab<4.5,>=4.4.0rc0->notebook)
(2.2.5)
Requirement already satisfied: setuptools>=41.1.0 in
./.conda/lib/python3.11/site-packages (from jupyterlab<4.5,>=4.4.0rc0->notebook)
(75.8.0)
Requirement already satisfied: babel>=2.10 in ./.conda/lib/python3.11/site-
packages (from jupyterlab-server<3,>=2.27.1->notebook) (2.17.0)
Requirement already satisfied: json5>=0.9.0 in ./.conda/lib/python3.11/site-
packages (from jupyterlab-server<3,>=2.27.1->notebook) (0.12.0)
Requirement already satisfied: requests>=2.31 in ./.conda/lib/python3.11/site-
packages (from jupyterlab-server<3,>=2.27.1->notebook) (2.32.3)
Requirement already satisfied: six>=1.5 in ./.conda/lib/python3.11/site-packages
(from python-dateutil>=2.8.2->pandas) (1.17.0)
Requirement already satisfied: wrapt in ./.conda/lib/python3.11/site-packages
(from smart-open>=1.8.1->gensim) (1.17.2)
Requirement already satisfied: idna>=2.8 in ./.conda/lib/python3.11/site-
packages (from anyio>=3.1.0->jupyter-server<3,>=2.4.0->notebook) (3.10)
Requirement already satisfied: sniffio>=1.1 in ./.conda/lib/python3.11/site-
packages (from anyio>=3.1.0->jupyter-server<3,>=2.4.0->notebook) (1.3.1)
Requirement already satisfied: argon2-cffi-bindings in
./.conda/lib/python3.11/site-packages (from argon2-cffi>=21.1->jupyter-
server<3,>=2.4.0->notebook) (21.2.0)
Requirement already satisfied: certifi in ./.conda/lib/python3.11/site-packages
(from httpx>=0.25.0->jupyterlab<4.5,>=4.4.0rc0->notebook) (2025.1.31)
Requirement already satisfied: httpcore==1.* in ./.conda/lib/python3.11/site-
packages (from httpx>=0.25.0->jupyterlab<4.5,>=4.4.0rc0->notebook) (1.0.8)
Requirement already satisfied: h11<0.15,>=0.13 in ./.conda/lib/python3.11/site-
packages (from
httpcore==1.*->httpx>=0.25.0->jupyterlab<4.5,>=4.4.0rc0->notebook) (0.14.0)
Requirement already satisfied: appnope in ./.conda/lib/python3.11/site-packages
(from ipykernel>=6.5.0->jupyterlab<4.5,>=4.4.0rc0->notebook) (0.1.2)
Requirement already satisfied: debugpy>=1.6.5 in ./.conda/lib/python3.11/site-
packages (from ipykernel>=6.5.0->jupyterlab<4.5,>=4.4.0rc0->notebook) (1.8.11)
Requirement already satisfied: nest-asyncio in ./.conda/lib/python3.11/site-
packages (from ipykernel>=6.5.0->jupyterlab<4.5,>=4.4.0rc0->notebook) (1.6.0)
Requirement already satisfied: psutil in ./.conda/lib/python3.11/site-packages
```

```
(from ipykernel>=6.5.0->jupyterlab<4.5,>=4.4.0rc0->notebook) (5.9.0)
Requirement already satisfied: parso<0.9.0,>=0.8.4 in
./.conda/lib/python3.11/site-packages (from
jedi>=0.16->ipython>=6.1.0->ipywidgets) (0.8.4)
Requirement already satisfied: MarkupSafe>=2.0 in ./.conda/lib/python3.11/site-
packages (from jinja2>=3.0.3->jupyter-server<3,>=2.4.0->notebook) (3.0.2)
Requirement already satisfied: python-json-logger>=2.0.4 in
./.conda/lib/python3.11/site-packages (from jupyter-events>=0.11.0->jupyter-
server<3,>=2.4.0->notebook) (3.3.0)
Requirement already satisfied: pyyaml>=5.3 in ./.conda/lib/python3.11/site-
packages (from jupyter-events>=0.11.0->jupyter-server<3,>=2.4.0->notebook)
(6.0.2)
Requirement already satisfied: rfc3339-validator in
./.conda/lib/python3.11/site-packages (from jupyter-events>=0.11.0->jupyter-
server<3,>=2.4.0->notebook) (0.1.4)
Requirement already satisfied: rfc3986-validator>=0.1.1 in
./.conda/lib/python3.11/site-packages (from jupyter-events>=0.11.0->jupyter-
server<3,>=2.4.0->notebook) (0.1.1)
Requirement already satisfied: beautifulsoup4 in ./.conda/lib/python3.11/site-
packages (from nbconvert>=6.4.4->jupyter-server<3,>=2.4.0->notebook) (4.13.3)
Requirement already satisfied: bleach!=5.0.0 in ./.conda/lib/python3.11/site-
packages (from bleach[css]!=5.0.0->nbconvert>=6.4.4->jupyter-
server<3,>=2.4.0->notebook) (6.2.0)
Requirement already satisfied: defusedxml in ./.conda/lib/python3.11/site-
packages (from nbconvert>=6.4.4->jupyter-server<3,>=2.4.0->notebook) (0.7.1)
Requirement already satisfied: jupyterlab-pygments in
./.conda/lib/python3.11/site-packages (from nbconvert>=6.4.4->jupyter-
server<3,>=2.4.0->notebook) (0.3.0)
Requirement already satisfied: mistune<4,>=2.0.3 in
./.conda/lib/python3.11/site-packages (from nbconvert>=6.4.4->jupyter-
server<3,>=2.4.0->notebook) (3.1.3)
Requirement already satisfied: nbclient>=0.5.0 in ./.conda/lib/python3.11/site-
packages (from nbconvert>=6.4.4->jupyter-server<3,>=2.4.0->notebook) (0.10.2)
Requirement already satisfied: pandocfilters>=1.4.1 in
./.conda/lib/python3.11/site-packages (from nbconvert>=6.4.4->jupyter-
server<3,>=2.4.0->notebook) (1.5.1)
Requirement already satisfied: ptyprocess>=0.5 in ./.conda/lib/python3.11/site-
packages (from pexpect>4.3->ipython>=6.1.0->ipywidgets) (0.7.0)
Requirement already satisfied: wcwidth in ./.conda/lib/python3.11/site-packages
(from prompt-toolkit<3.1.0,>=3.0.41->ipython>=6.1.0->ipywidgets) (0.2.5)
Requirement already satisfied: charset-normalizer<4,>=2 in
./.conda/lib/python3.11/site-packages (from requests>=2.31->jupyterlab-
server<3,>=2.27.1->notebook) (3.4.1)
Requirement already satisfied: urllib3<3,>=1.21.1 in
./.conda/lib/python3.11/site-packages (from requests>=2.31->jupyterlab-
server<3,>=2.27.1->notebook) (2.4.0)
Requirement already satisfied: executing in ./.conda/lib/python3.11/site-
packages (from stack-data->ipython>=6.1.0->ipywidgets) (0.8.3)
```

```
Requirement already satisfied: asttokens in ./.conda/lib/python3.11/site-
packages (from stack-data->ipython>=6.1.0->ipywidgets) (3.0.0)
Requirement already satisfied: pure-eval in ./.conda/lib/python3.11/site-
packages (from stack-data->ipython>=6.1.0->ipywidgets) (0.2.2)
Requirement already satisfied: webencodings in ./.conda/lib/python3.11/site-
packages (from bleach!=5.0.0->bleach[css]!=5.0.0->nbconvert>=6.4.4->jupyter-
server<3,>=2.4.0->notebook) (0.5.1)
Requirement already satisfied: tinycss2<1.5,>=1.1.0 in
./.conda/lib/python3.11/site-packages (from
bleach[css]!=5.0.0->nbconvert>=6.4.4->jupyter-server<3,>=2.4.0->notebook)
(1.4.0)
Requirement already satisfied: fqdn in ./.conda/lib/python3.11/site-packages
(from jsonschema[format-nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-
server<3,>=2.4.0->notebook) (1.5.1)
Requirement already satisfied: isoduration in ./.conda/lib/python3.11/site-
packages (from jsonschema[format-nongpl]>=4.18.0->jupyter-
events>=0.11.0->jupyter-server<3,>=2.4.0->notebook) (20.11.0)
Requirement already satisfied: jsonpointer>1.13 in ./.conda/lib/python3.11/site-
packages (from jsonschema[format-nongpl]>=4.18.0->jupyter-
events>=0.11.0->jupyter-server<3,>=2.4.0->notebook) (3.0.0)
Requirement already satisfied: uri-template in ./.conda/lib/python3.11/site-
packages (from jsonschema[format-nongpl]>=4.18.0->jupyter-
events>=0.11.0->jupyter-server<3,>=2.4.0->notebook) (1.3.0)
Requirement already satisfied: webcolors>=24.6.0 in
./.conda/lib/python3.11/site-packages (from jsonschema[format-
nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-server<3,>=2.4.0->notebook)
(24.11.1)
Requirement already satisfied: cffi>=1.0.1 in ./.conda/lib/python3.11/site-
packages (from argon2-cffi-bindings->argon2-cffi>=21.1->jupyter-
server<3,>=2.4.0->notebook) (1.17.1)
Requirement already satisfied: soupsieve>1.2 in ./.conda/lib/python3.11/site-
packages (from beautifulsoup4->nbconvert>=6.4.4->jupyter-
server<3,>=2.4.0->notebook) (2.6)
Requirement already satisfied: pycparser in ./.conda/lib/python3.11/site-
packages (from cffi>=1.0.1->argon2-cffi-bindings->argon2-cffi>=21.1->jupyter-
server<3,>=2.4.0->notebook) (2.22)
Requirement already satisfied: arrow>=0.15.0 in ./.conda/lib/python3.11/site-
packages (from isoduration->jsonschema[format-nongpl]>=4.18.0->jupyter-
events>=0.11.0->jupyter-server<3,>=2.4.0->notebook) (1.3.0)
Requirement already satisfied: types-python-dateutil>=2.8.10 in
./.conda/lib/python3.11/site-packages (from
arrow>=0.15.0->isoduration->jsonschema[format-nongpl]>=4.18.0->jupyter-
events>=0.11.0->jupyter-server<3,>=2.4.0->notebook) (2.9.0.20241206)
Note: you may need to restart the kernel to use updated packages.
```

[ ]:
```
import gensim.downloader as api
from sklearn.manifold import TSNE
```

```
import numpy as np
import plotly.express as px
import pandas as pd
import plotly.io as pio
```

1.2.1 Φ Word2Vec GloVe

X embeddings: - word2vec-google-news-300: 300- Google - glove-wiki-gigaword-300: 300- Wikipedia Gigaword corpus

H gensim.downloader.

[3]:
```
# Word2Vec Google News (300 )
w2v_model = api.load("word2vec-google-news-300")

# GloVe Wiki Gigaword (300 )
glove_model = api.load("glove-wiki-gigaword-300")
```

1.3 B Σ

O .

- E , print_similar_words :
  - top-N ,
  - : ,
  - .

[ ]:
```
# O Σ Similar Words
def print_similar_words(model, word, topn=10):
    try:
        # Λ top-N
        similar = model.most_similar(word, topn=topn)
        print(f"Top {topn} most similar words to '{word}':")

        # E
        for w, score in similar:
            print(f"  {w:<15} {score:.4f}")
        return set(w for w, _ in similar)

        # X
    except KeyError:
        print(f"Word '{word}' not in vocabulary.")
        return set()
```

1.3.1 E 1 – Λ O (Word Similarity)

Σ , 'car', 'jaguar', 'Jaguar' 'facebook' word embeddings:

- **Word2Vec (Google News, 300)**
- **GloVe (Wikipedia + Gigaword, 300)**

Γ , 10 **(cosine similarity)**.
Σ , .

[5]:
```
# E          4
target_words = ['car', 'jaguar', 'Jaguar', 'facebook']

print("\n=== Word2Vec Results ===")
w2v_results = {word: print_similar_words(w2v_model, word) for word in␣
 ↪target_words}

print("\n=== GloVe Results ===")
glove_results = {word: print_similar_words(glove_model, word) for word in␣
 ↪target_words}
```

```
=== Word2Vec Results ===
Top 10 most similar words to 'car':
  vehicle         0.7821
  cars            0.7424
  SUV             0.7161
  minivan         0.6907
  truck           0.6736
  Car             0.6678
  Ford_Focus      0.6673
  Honda_Civic     0.6627
  Jeep            0.6511
  pickup_truck    0.6441
Top 10 most similar words to 'jaguar':
  jaguars         0.6738
  Macho_B         0.6313
  panther         0.6086
  lynx            0.5815
  rhino           0.5754
  lizard          0.5607
  tapir           0.5563
  tiger           0.5529
  leopard         0.5473
  Florida_panther 0.5464
Top 10 most similar words to 'Jaguar':
  Land_Rover      0.6484
  Aston_Martin    0.6437
  Mercedes        0.6420
  Porsche         0.6233
  BMW             0.6055
```

```
  Bentley_Arnage  0.6040
  XF_sedan        0.5996
  Audi            0.5976
  Jaguar_XF       0.5951
  XJ_saloon       0.5942
Top 10 most similar words to 'facebook':
  Facebook        0.7564
  FaceBook        0.7077
  twitter         0.6989
  myspace         0.6942
  Twitter         0.6642
  twitter_facebook 0.6572
  Facebook.com    0.6530
  myspace_facebook 0.6371
  facebook_twitter 0.6368
  linkedin        0.6357

=== GloVe Results ===
Top 10 most similar words to 'car':
  cars            0.7827
  vehicle         0.7655
  truck           0.7351
  driver          0.7115
  driving         0.6442
  vehicles        0.6328
  motorcycle      0.6023
  automobile      0.5956
  parked          0.5910
  drivers         0.5778
Top 10 most similar words to 'jaguar':
  rover           0.5931
  bmw             0.5415
  mercedes        0.5256
  sepecat         0.5030
  mustang         0.4987
  lexus           0.4845
  volvo           0.4829
  cosworth        0.4809
  xk              0.4764
  maserati        0.4757
Word 'Jaguar' not in vocabulary.
Top 10 most similar words to 'facebook':
  twitter         0.8350
  myspace         0.8056
  youtube         0.7292
  blog            0.6404
  linkedin        0.6333
  google          0.6268
```

```
website         0.6157
web             0.6143
blogs           0.6064
networking      0.6047
```

[6]:
```
# Υ      Κ    Λ
for word in target_words:
    common = w2v_results[word].intersection(glove_results[word])
    print(f"\nCommon similar words for '{word}': {len(common)}")
    print(common)
```

```
Common similar words for 'car': 3
{'vehicle', 'cars', 'truck'}

Common similar words for 'jaguar': 0
set()

Common similar words for 'Jaguar': 0
set()

Common similar words for 'facebook': 3
{'myspace', 'linkedin', 'twitter'}
```

1.3.2 Π Σ (E 1)

H Word2Vec GloVe 'car', 'jaguar', 'Jaguar' 'facebook' :

- H 'jaguar' Word2Vec panther, lynx, leopard, . ., GloVe mercedes, mustang, land_rover. A context .
- H 'Jaguar' **J**:
  - ϒ Word2Vec .
  - **Δ** GloVe, .
- Γ 'facebook', GloVe linkedin, twitter, myspace, Word2Vec twitter_facebook, facebook_twitter, :
  - T Word2Vec **underscore** , corpus Google.
  - T GloVe **(normalization)**.
- ϒ **0** 'jaguar' 'Jaguar', **3** 'car' 'facebook', ( ) , (. . 'jaguar') .

1.3.3 E 2 – E Word Embeddings Λ E

Σ , 4 T N Π :
'AI', 'Python', 'Analytics', 'Machine'.

A E 1: - E 10 , Word2Vec
GloVe. - ϒ .

Σ embeddings « »

[7]:
```
# E     Δ         E    2
custom_words = ['AI', 'Python', 'Analytics', 'Machine']

print("\n=== Word2Vec Results (Custom Words) ===")
w2v_custom = {word: print_similar_words(w2v_model, word) for word in␣
 ↪custom_words}

print("\n=== GloVe Results (Custom Words) ===")
glove_custom = {word: print_similar_words(glove_model, word) for word in␣
 ↪custom_words}

print("\n=== Common Similar Words Between Word2Vec and GloVe ===")
for word in custom_words:
    common = w2v_custom[word].intersection(glove_custom[word])
    print(f"\nCommon similar words for '{word}': {len(common)}")
    print(common)
```

```
=== Word2Vec Results (Custom Words) ===
Top 10 most similar words to 'AI':
  Steven_Spielberg_Artificial_Intelligence 0.5576
  Index_MDE_##/###/#### 0.5415
  Enemy_AI        0.5256
  Ace_Combat_Zero 0.5227
  DOA4            0.5183
  mechs           0.5137
  mech            0.5078
  playstyle       0.5073
  AI_bots         0.5051
  deathmatch_mode 0.5046
Top 10 most similar words to 'Python':
  Jython          0.6153
  Perl_Python     0.5711
  IronPython      0.5705
  scripting_languages 0.5695
  PHP_Perl        0.5688
  Java_Python     0.5681
  PHP             0.5661
  Python_Ruby     0.5632
```

```
  Visual_Basic    0.5603
  Perl            0.5531
Top 10 most similar words to 'Analytics':
  analytics       0.6786
  Text_Analytics  0.5867
  predictive_analytics 0.5806
  Optimization    0.5766
  analytic_tools  0.5713
  Interwoven_Segmentation 0.5705
  TRX_Travel      0.5701
  Metrics         0.5599
  Chmura_Economics 0.5578
  Performance_Dashboards 0.5563
Top 10 most similar words to 'Machine':
  Machine_Audioslave 0.6385
  Machines        0.6147
  Machine_Killing 0.5741
  Machine_humbles 0.5441
  machine         0.5064
  Manufacturing_ISIN_AT########## 0.5063
  GigaPan_Time    0.4952
  Pearl_Jam_Rage_Against 0.4931
  Tool            0.4914
  ElectraTherm_Green 0.4909

=== GloVe Results (Custom Words) ===
Word 'AI' not in vocabulary.
Word 'Python' not in vocabulary.
Word 'Analytics' not in vocabulary.
Word 'Machine' not in vocabulary.

=== Common Similar Words Between Word2Vec and GloVe ===

Common similar words for 'AI': 0
set()

Common similar words for 'Python': 0
set()

Common similar words for 'Analytics': 0
set()

Common similar words for 'Machine': 0
set()
```

1.3.4 Π Σ (E 2)

H 'AI', 'Python', 'Analytics' 'Machine' :

- K **GloVe** (case sensitivity preprocessing), :
  – corpus ,
  – (. .).
- T Word2Vec, , . I :
  – H `'Python'` .
  – A Word2Vec (Google News) .
- H (`'jaguar'`) :
  – T Word2Vec (. . `'jaguar'` `'Python'`).
  – T GloVe « » , ,
  .

Σ , embedding .

1.3.5 E 3 – E Σ Π

Σ `'student'` embeddings (Word2Vec GloVe).
Σ « » (context) .

A 10 `'student'` .
E , : - (`university`, `campus`, `semester`, . .) - (`teacher`, `homework`, `classroom`, . .)

« » `'student'`.

[8]:
```
#                student

print("\n--- Word2Vec - Top 10 similar to 'student' ---")
w2v_student = print_similar_words(w2v_model, 'student')

print("\n--- GloVe - Top 10 similar to 'student' ---")
glove_student = print_similar_words(glove_model, 'student')
```

```
--- Word2Vec - Top 10 similar to 'student' ---
Top 10 most similar words to 'student':
  students        0.7295
  Student         0.6707
  teacher         0.6301
  stu_dent        0.6241
  faculty         0.6087
  school          0.6056
  undergraduate   0.6020
  university      0.6005
  undergraduates  0.5756
  semester        0.5738
```

```
--- GloVe - Top 10 similar to 'student' ---
Top 10 most similar words to 'student':
  students        0.7691
  teacher         0.6874
  graduate        0.6738
  school          0.6131
  college         0.6090
  undergraduate   0.6044
  faculty         0.5999
  university      0.5971
  academic        0.5810
  campus          0.5768
```

[9]:
```python
# E                 "University"

university_related = {
    'college', 'university', 'campus', 'undergraduate', 'graduate',
    'professor', 'semester', 'dorm', 'sophomore', 'finals', 'academic'
}

print("\n--- Word2Vec (excluding university context) ---")
w2v_no_uni = [w for w in w2v_model.most_similar('student', topn=20) if w[0].
 ↪lower() not in university_related][:10]
for w, score in w2v_no_uni:
    print(f"{w:<15} {score:.4f}")

print("\n--- GloVe (excluding university context) ---")
glove_no_uni = [w for w in glove_model.most_similar('student', topn=20) if w[0].
 ↪lower() not in university_related][:10]
for w, score in glove_no_uni:
    print(f"{w:<15} {score:.4f}")
```

```
--- Word2Vec (excluding university context) ---
students        0.7295
Student         0.6707
teacher         0.6301
stu_dent        0.6241
faculty         0.6087
school          0.6056
undergraduates  0.5756
classmates      0.5528
Students        0.5501
undergrad       0.5432

--- GloVe (excluding university context) ---
students        0.7691
```

```
teacher         0.6874
school          0.6131
faculty         0.5999
teachers        0.5537
education       0.5337
enrolled        0.5298
teaching        0.5292
colleges        0.5042
harvard         0.5041
```

[10]:
```python
# E                 "School"
school_related = {
    'school', 'teacher', 'classroom', 'homework', 'principal', 'elementary',
    'middle', 'highschool', 'pupil', 'curriculum', 'grade'
}

print("\n--- Word2Vec (excluding school context) ---")
w2v_no_school = [w for w in w2v_model.most_similar('student', topn=20) if w[0].
 ↪lower() not in school_related][:10]
for w, score in w2v_no_school:
    print(f"{w:<15} {score:.4f}")

print("\n--- GloVe (excluding school context) ---")
glove_no_school = [w for w in glove_model.most_similar('student', topn=20) if␣
 ↪w[0].lower() not in school_related][:10]
for w, score in glove_no_school:
    print(f"{w:<15} {score:.4f}")
```

```
--- Word2Vec (excluding school context) ---
students        0.7295
Student         0.6707
stu_dent        0.6241
faculty         0.6087
undergraduate   0.6020
university      0.6005
undergraduates  0.5756
semester        0.5738
campus          0.5629
classmates      0.5528

--- GloVe (excluding school context) ---
students        0.7691
graduate        0.6738
college         0.6090
undergraduate   0.6044
faculty         0.5999
university      0.5971
```

```
academic        0.5810
campus          0.5768
teachers        0.5537
education       0.5337
```

1.3.6 Π Σ (E 3)

- K , :
  'students', 'teacher', 'school', 'university', 'semester', 'undergraduate'.
- **H** / Word2Vec ('Student' vs 'student'),
  context, .
- Ø **“university”**, :
  – Σ 'harvard', 'graduates', 'courses',
  , .
  – T “ ” , .
- Ø **“school” context**, :
  .. 'semester', 'undergraduates', 'university', 'campus' .

T :

M .
A , fine-tuning, embeddings
.

1.3.7 E 4 – A Λ (Word Analogies)

Σ Word2Vec GloVe
.

O : > **A − B + C = ?**

Γ , : > 'king' - 'man' + 'woman'
'queen'.

O : - / (.. 'doctor' - 'father' + 'mother') -
(.. 'France' - 'Paris' + 'Tokyo') - (.. 'swimming' - 'walking' + 'walked')

H most_similar(positive=[...], negative=[...]) .

[11]:
```
# O    Σ     A
def print_analogy_result(model, positive, negative, topn=5):
    try:
        result = model.most_similar(positive=positive, negative=negative,␣
↪topn=topn)
        print(f"\nAnalogy: {positive[0]} - {negative[0]} + {positive[1]} = ?")
        for w, score in result:
            if w not in positive + negative:
```

```
                print(f"  {w:<15} {score:.4f}")
    except KeyError as e:
        print(f"Word not in vocabulary: {e}")
```

[12]:
```
analogies = [
    (["king", "woman"], ["man"]),
    (["france", "tokyo"], ["paris"]),
    (["trees", "grapes"], ["apples"]),
    (["swimming", "walked"], ["walking"]),
    (["doctor", "mother"], ["father"])
]


print("=== Word2Vec - A             ===")
for pos, neg in analogies:
    print_analogy_result(w2v_model, positive=pos, negative=neg)

print("\n=== GloVe - A             ===")
for pos, neg in analogies:
    print_analogy_result(glove_model, positive=pos, negative=neg)
```

```
=== Word2Vec - A             ===

Analogy: king - man + woman = ?
  queen           0.7118
  monarch         0.6190
  princess        0.5902
  crown_prince    0.5499
  prince          0.5377

Analogy: king - man + woman = ?
  queen           0.7118
  monarch         0.6190
  princess        0.5902
  crown_prince    0.5499
  prince          0.5377

Analogy: france - paris + tokyo = ?
  japan           0.5508
  hong_kong       0.5012
  japanese        0.4837
  seoul           0.4790
  germany         0.4736

Analogy: trees - apples + grapes = ?
  oak_trees       0.6750
  vines           0.6702
  pine_trees      0.6573
```

```
  oaks            0.6505
  tree            0.6358

Analogy: swimming - walking + walked = ?
  swam            0.6926
  swim            0.6725
  swimmers        0.5923
  swum            0.5857
  Swimming        0.5806

Analogy: doctor - father + mother = ?
  nurse           0.7128
  doctors         0.6593
  gynecologist    0.6454
  physician       0.6408
  nurse_practitioner 0.6387

=== GloVe - A             ===

Analogy: king - man + woman = ?
  queen           0.6713
  princess        0.5433
  throne          0.5386
  monarch         0.5348
  daughter        0.4980

Analogy: france - paris + tokyo = ?
  japan           0.8017
  japanese        0.6111
  korea           0.5508
  yen             0.4853
  taiwan          0.4487

Analogy: trees - apples + grapes = ?
  vines           0.5909
  tree            0.5843
  planted         0.5468
  forests         0.5134
  grape           0.4985

Analogy: swimming - walking + walked = ?
  swam            0.4978
  swimmers        0.4852
  pool            0.4667
  swimmer         0.4602
  athletics       0.4583

Analogy: doctor - father + mother = ?
```

```
nurse           0.6570
doctors         0.6172
woman           0.5800
patient         0.5768
pregnant        0.5368
```

1.3.8 Π Σ (E 4)

- K Word2Vec GloVe , :
  - `'king' - 'man' + 'woman'` `queen`
  - `'swimming' - 'walking' + 'walked'` `swam`
    → (,).
- Ω , **(bias)**:
  - `'doctor' - 'father' + 'mother'` `nurse`
    → , “ ” “ ”,
    .
- E :
  - `'trees' - 'apples' + 'grapes'`
    - Word2Vec → `oak_trees` (0.67)
    - GloVe → `vines`, (0.59)
- Σ `'France' - 'Paris' + 'Tokyo'`:
  - Word2Vec `'Japan'` **(0.55)**
  - GloVe `'Japan'` **(0.80)**
    → T GloVe « », .

Γ :

T ,
:

- “ ” (vs),
- corpus .

1.3.9 E 5 – A Π Λ

Σ , , « » embeddings.
O « » :

- (. . `"Europe" - "Greece" + "Washington"`)
- (. . `"War" - "Peace" + "Love"`)
- (. . `"Olympiacos" - "Piraeus" + "Rome"`)

O E 4, Word2Vec GloVe.

[13]:
```
custom_analogies = [
    (["Olympiacos", "Rome"], ["Pireaus"]),
    (["europe", "washington"], ["greece"]),
    (["war", "love"] , ["peace"])
]


print("=== Word2Vec - Custom Analogies ===")
for pos, neg in custom_analogies:
    print_analogy_result(w2v_model, positive=pos, negative=neg)

print("\n=== GloVe - Custom Analogies ===")
for pos, neg in custom_analogies:
    print_analogy_result(glove_model, positive=pos, negative=neg)
```

```
=== Word2Vec - Custom Analogies ===

Analogy: Olympiacos - Pireaus + Rome = ?
  Milan           0.4649
  Olympiakos      0.4437
  Juve            0.4412
  Juventus        0.4361
  AC_Milan        0.4295

Analogy: europe - greece + washington = ?
  america         0.5460
  usa             0.5049
  florida         0.4940
  michigan        0.4937
  obama           0.4796

Analogy: war - peace + love = ?
  hate            0.4494
  loved           0.4454
  wars            0.4423
  adore           0.4307
  loves           0.4253

=== GloVe - Custom Analogies ===
Word not in vocabulary: "Key 'Olympiacos' not present in vocabulary"

Analogy: europe - greece + washington = ?
  d.c.            0.5270
  america         0.4957
  states          0.4930
  united          0.4871
  u.s.            0.4851
```

```
Analogy: war - peace + love = ?
  tale            0.4804
  romance         0.4726
  movie           0.4645
  ?               0.4595
  passion         0.4572
```

1.3.10 Π Σ (E 5)

- K 'Olympiacos' **GloVe**, corpus preprocessing.
  A GloVe Word2Vec.
- T Word2Vec « » 'Olympiacos' ,
  'Milan', 'Juventus', 'AC_Milan'.
  A , .
- Σ 'Europe' - 'Greece' + 'Washington':
  - T Word2Vec 'america', .
  - T GloVe, , 'd.c.', 'america', 'states' . .
- H Word2Vec 'War' - 'Peace' + 'Love',
  'Hate'.
  T : : ,
- T GloVe, , 'romance', 'tale', 'movie', ,
  .

Γ Σ :

T **Word2Vec** « » ,
.

T **GloVe** « », ,
.

1.3.11 E 6 – O Word Embeddings t-SNE

Σ **t-SNE** (t-distributed Stochastic Neighbor Embedding)
300- GloVe 2 , .

E : - (. . school, student, homework) - (. . job,
manager, employee)

X **Plotly** : - , -
legend, - **zoom** **pan** .

O embeddings (context).

[14]:
```
# Λ
words = [
```

```
    'assignment', 'exam', 'career', 'classroom', 'colleague', 'college',␣
 ↪'coworker',
    'curriculum', 'degree', 'employee', 'employment', 'grade', 'homework',␣
 ↪'job',
    'learning', 'lecture', 'lesson', 'major', 'manager', 'occupation', 'office',
    'position', 'profession', 'school', 'student', 'subject', 'supervisor',␣
 ↪'teacher',
    'teaching', 'test', 'trade', 'university', 'vocation', 'workplace'
]
```

[15]:
```
# A              GloVe
vectors = []
valid_words = []

for word in words:
    try:
        vectors.append(glove_model[word])
        valid_words.append(word)
    except KeyError:
        print(f"'{word}' not found in vocabulary.")
```

[16]:
```
# E     t-SNE
tsne = TSNE(n_components=2, random_state=42, perplexity=5)
vectors_2d = tsne.fit_transform(np.array(vectors))
```

[32]:
```
pio.renderers.default = 'notebook'

df = pd.DataFrame(vectors_2d, columns=['x', 'y'])
df['word'] = valid_words

fig = px.scatter(
    df, x='x', y='y', text='word',
    color='word',              #                  →          legend
    hover_name='word',
    title="<b> t-SNE of GloVe Word Embeddings </b>",
    width=1280, height=720
)

# M              label (  hover    legend intact)
fig.update_traces(
    marker=dict(size=10),
    textposition="top center",
    mode='markers+text',
)

fig.update_layout(
    legend_title_text='Word',
```

```
    showlegend=True,
    xaxis_title='',
    yaxis_title=''
)

fig.show()
```

1.3.12 Π Σ (E 6)

H « » (clusters) :

- O **coworker, colleague, employee, manager, workplace** , : .
- A , **classroom, curriculum, teaching, learning, school** , .
- Π **exam test** , , context corpus.
- Π , **profession occupation** , . A :
  - corpus,
  - context (. . **profession** / , **occupation** /).
- T , **trade major** :
  - , . . **major** , , . .

Σ :

H t-SNE GloVe embeddings .

M clusters, outliers , Plotly .