DATA.ML.200 Pattern Recognition and Machine Learning

TAU Signal Processing
Exercise Week 1: Machine learning fundamentals

This exercise you need to do without any help from anyone. The exercise verifies that your programming skills and machine learning background are sufficient for the course. This exercise is **mandatory** to pass the course.

1. **`python`** Nearest neighbor classifier (20 points)

   a) Write code to load and vectorize the 20 Newsgroups dataset (available in the lecture notebook)

   b) Classify the test samples using a *simple baseline* method for classification. You may implement the baseline yourself or use the Sklearn `DummyClassifier` module using the 'most frequent class' strategy.

   c) Add to your code functionality to measure the computation time of classifying all test samples. Suitable functions can be found from the Python 'timeit' library (for example: *.default_timer()*).

   d) Print the classification accuracy and computation time for the baseline.

   e) Implement a straightforward nearest neigbor classifier that classifies each test sample one by one. At first, Euclidean distances are computed to all training samples, and then, class label of the smallest distance sample is used as a prediction.

   f) It may turn out that your own NN classifier is slow. In that case, use only a portion of the test samples, and use them to estimate the total computation time.

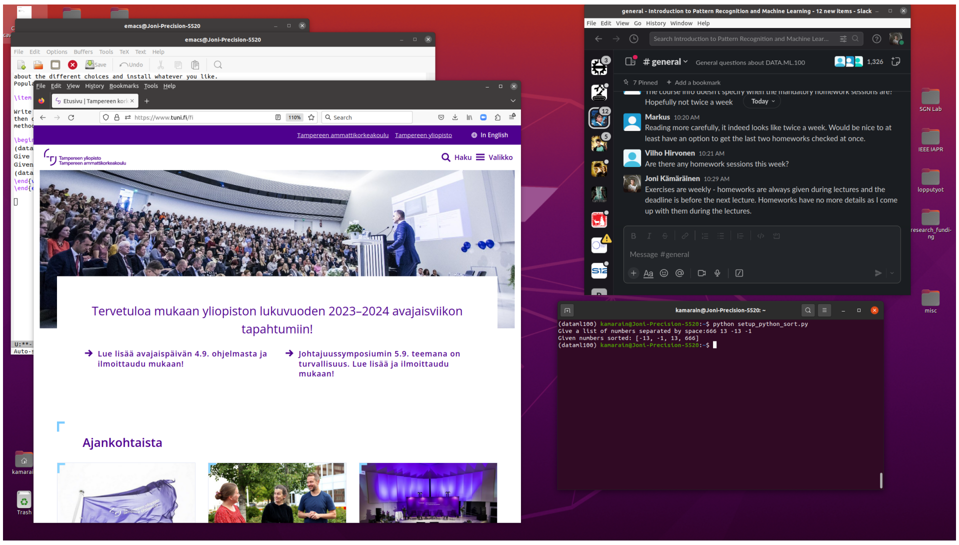
      Print the number of test samples used, estimated total computation time, and the classification accuracy.

   g) Replace your NN classifier with the Sklearn *NearestNeighbor* classifier.

   h) Print the computation time and classification accuracy for the Sklearn NN.

      Return the following items:

      - Python code: onenn.py that does all above.
      - PNG image: your full desktop screenshot that includes a terminal where the python file is executed and it prints all above numbers: onenn_screenshot.png (see the example)



emacs@Joni-Precision-5520
File Edit Options Buffers Tools TeX Text Help
Save
Undo
File Edit View History Bookmarks Tools Help
Etusivu | Tampereen kork
https://www.tuni.fi/fi
110%
Search
Tampereen ammattikorkeakoulu
Tampereen yliopisto
In English
Tampereen yliopisto
Tampereen ammattikorkeakoulu
Haku
Valikko
general - Introduction to Pattern Recognition and Machine Learning - 12 new items - Slack
File Edit View Go History Window Help
Search Introduction to Pattern Recognition and Machine Lear...
# general
General questions about DATA.ML.100
1,326
7 Pinned
Add a bookmark
Today
Message #general
SGN Lab
IEEE IAPR
lopputyot
research_funding
misc
kamarain@Joni-Precision-5520: ~
Trash

about the different choices and install whatever you like.

Tervetuloa mukaan yliopiston lukuvuoden 2023–2024 avajaisviikon tapahtumiin!

→ Lue lisää avajaispäivän 4.9. ohjelmasta ja ilmoittaudu mukaan!

→ Johtajuussymposiumin 5.9. teemana on turvallisuus. Lue lisää ja ilmoittaudu mukaan!

Ajankohtaista

Hopefully not twice a week

Markus 10:20 AM
Reading more carefully, it indeed looks like twice a week. Would be nice to at least have an option to get the last two homeworks checked at once.

Vilho Hirvonen 10:21 AM
Are there any homework sessions this week?

Joni Kämäräinen 10:29 AM
Exercises are weekly - homeworks are always given during lectures and the deadline is before the next lecture. Homeworks have no more details as I come up with them during the lectures.

```
(dataml100) kamarain@Joni-Precision-5520:~$ python setup_python_sort.py
Give a list of numbers separated by space:666 13 -13 -1
Given numbers sorted: [-13, -1, 13, 666]
(dataml100) kamarain@Joni-Precision-5520:~$
```

Figure 1: Example of the required screenshot