

Preferred Networks Internship 2022 Thematic Task

We question the understanding of the internship's thematic expertise. Select a question corresponding to the theme of your first choice from the list below and submit your answer.

If the task is in a report format, please submit the answer in an A4 PDF with the file name `survey.pdf`. Please follow the format specified in each question regarding page limits and format. Unless otherwise specified, the number of pages is assumed to be no more than two including the reference section, and the format is assumed to be free.

If the task is not in a report format, please submit the answer following the format specified in each question. If the specified format is `zip`, please submit the answer with the file name `survey.zip`.

On the following themes, there are no thematic tasks. Please solve only the coding task.

- JE03. Development and operation of a web system for Matlantis which is a general-purpose atomic-level simulator
- JE05. Development of library and/or user interface to accelerate drug discovery research

Notice

Please do the task yourself. Do not share or discuss this task with anyone, including other applicants. **Do not upload your solution and/or problem description to a public repository on GitHub or social networking sites.** (After the screening period finished, we may publish the problem statement on GitHub. If so, you may share your solution with others.) If we find evidence of leakage, the applicant will be disqualified. If an applicant allows another applicant to copy answers, both applicants will be disqualified.

We expect the coding task and the thematic task to take up to two days. You can submit your completed work without solving all of the problems. Please do your best without neglecting your coursework.

Changelog

- Apr 27th, 2022 : Initial version

JE01. Development of Framework and Library for Deploying Deep Learning Models in Real World

Please read the pdf file in [je01](#) ([je01/README-en.pdf](#)).

JE02. Development and optimization of storage systems for deep learning and simulation

Please describe either or both processes in detail as much as you can, based on performance perspective. File system and the storage device may be whatever you like, but a specific example in Linux is required.

- In read(2), the process of transferring binary data in the storage media to the RAM in user land, from the system call until the control comes back to the user program
- In write(2), the process of transferring the data on the user land RAM to the storage media, until the control comes back to the user program

JE03. Development and operation of a web system for Matlantis which is a general-purpose atomic-level simulator

(This theme has no thematic tasks.)

JE04. Optuna Development

(This theme's task is not open to the public.)

JE05. Development of library and/or user interface to accelerate drug discovery research

(This theme has no thematic tasks.)

JE06. Development of algorithms to calculate physical properties for Matlantis

Solve the following optimization problem. The function to be optimized is defined in the distributed jupyter notebook ([je06/JE06_task-en.ipynb](#)), so please refer to it for your solution.

If you find "JE19. Applied research and development of machine learning and atomic simulation on materials" easier to work on than this assignment, you may answer that assignment instead of this one.

Problem

Place 64 labeled points in 3-dimensional space. The positions of the points are assumed to be non-overlapping. The labels are integer values. For each integer from 0 to 7, there are 8 points with a label with a value equal to that integer. Consider a function that returns a scalar value for this set of points. In this case, consider the arrangement of the points such that the return value is as small as possible.

This assignment is modeled after the problem of estimating atomic structure. By considering each point as an atom and the scalar value as energy, the problem can be regarded as a problem of finding an atomic arrangement that is low in energy and stable.

Solution

Please submit a report (no more than one sheet of A4 paper in PDF format, or a little over a sheet if you cannot write on one sheet) describing the results (input pairs and their scores) and how you thought about them, as well as the source code of the program used to run the simulation. You may use the distributed Jupyter Notebook as-is for your program, or you may use a different format. For this assignment, you do not need to spend a lot of time and work down to the very last minute on your score. Instead, please consider readability when writing your program. Also, be sure to write in your report what you have considered in your optimization.

JE07. Development of Creative Tools

(This theme's task is not open to the public.)

JE08. Resolving the gap between training and test environments

Hypothetically, suppose that you were asked to become a reviewer for an international conference. Choose one paper from the following list, and write a review to the paper you selected. The followings are the steps we would like you to follow.

Step1. Choose a paper to review

Pick one paper from the following list.

- ICML 2021 <https://proceedings.mlr.press/v139/>

Step2. Write a review

Make sure to follow the subinstructions below:

- (1) Please write the Paper title, the Author, and the paper URL (e.g., <https://proceedings.mlr.press/v139/abdolshah21a.html>)
- (2) Please summarize the main claims of the paper you picked, as well as its significance. Please write as concisely as you can.
- (3) Please list three strong points of the paper.
- (4) Please list three weak points of the paper.
- (5) Please write your free opinions on this paper in detail, from as neutral a perspective as possible.

- You are not required to make a decision on this paper (e.g., reject, accept) nor to ask a question to the author.

Make sure to submit your "virtual review" in pdf format. The review is not to exceed 2 pages (font size >10pt).

JE09. Deep learning for tabular data with missing values

The task is given in the English language but applicants can answer either in Japanese or English. The maximum page length is **two pages** in A4 format (the minimum font size is 10 pts.). We only accept the submission in .pdf format. It is allowed to add figures and tables.

The purpose of this thematic task is to evaluate your familiarity with this topic and your overall research skill. There are 4 questions. We highly encourage you to provide some references (e.g., conference/journal papers) to support your answers.

Here, we provide a list of references that could be useful (it is allowed to also cite the papers that are not listed here):

- Van Buuren, S. (2018). Flexible imputation of missing data. CRC press. (free online access book: <https://stefvanbuuren.name/fimd/>)
- Yoon, J., Jordon, J., & Schaar, M. (2018, July). GAIN: Missing data imputation using generative adversarial nets. In International conference on machine learning (pp. 5689-5698). PMLR. (<https://proceedings.mlr.press/v80/yoon18a.html>)
- Nazabal, A., Olmos, P. M., Ghahramani, Z., & Valera, I. (2020). Handling incomplete heterogeneous data using vaes. Pattern Recognition, 107, 107501. (<https://arxiv.org/abs/1807.03653>)
- Gondara, L., & Wang, K. (2018, June). Mida: Multiple imputation using denoising autoencoders. In Pacific-Asia conference on knowledge discovery and data mining (pp. 260-272). Springer, Cham. (<https://arxiv.org/abs/1705.02737>)
- Tashiro, Y., Song, J., Song, Y., & Ermon, S. (2021). CSDI: Conditional Score-based Diffusion Models for Probabilistic Time Series Imputation. Advances in Neural Information Processing Systems, 34. (<https://arxiv.org/abs/2107.03502>)
- Wang, Z., Akande, O., Poulos, J., & Li, F. (2021). Are deep learning models superior for missing data imputation in large surveys? Evidence from an empirical comparison. arXiv preprint arXiv:2103.09316. (<https://arxiv.org/abs/2103.09316>)
- Hegde, H., Shimpi, N., Panny, A., Glurich, I., Christie, P., & Acharya, A. (2019). MICE vs PPCA: Missing data imputation in healthcare. Informatics in Medicine Unlocked, 17, 100275. (<https://www.sciencedirect.com/science/article/pii/S2352914819302783>)

Question 1

Suppose we want to train a machine learning model to predict urinary sugar level (尿糖), which has five different levels (1-5: discrete from normal to highly unsafe) from available health information of a person. In this problem, urinary sugar level 1 is considered normal. From level 2, it can be considered as abnormal.

The dataset for training the model consists of the health information of 10,000 people collected from several sources in a tabular format: (1) 1000 elderly people from nursing homes, (2) 3000 students from

universities, (3) 5000 patients from hospitals, and (4) 1000 players from soccer clubs.

For each person i , we are given the following information:

1. Feature vector: 100-dimensional feature vector x_i represents a person's health information: systolic blood pressure, diastolic blood pressure, height, weight, age, etc. Each feature can be either in a real-valued or categorical format. However, in our problem, some people may not have all 100 information we need, that is, there are missing values. For example, it is observed when collecting the data that 20% of patients in hospitals are not comfortable providing their age information. But all university students provide the age information. Another example we found is 80% of elderly people had taken a diabetes screening test before, while none of the university students took it. Nevertheless, it is guaranteed that everyone provides height, weight, blood pressure, blood sugar level, and cholesterol level.
2. Source label: $s_i \in \{\text{nursing homes, universities, hospitals, soccer clubs}\}$, which indicates where the data is collected from. It is guaranteed that s_i is not missing for any person i . Thus, it is possible to identify the data source for each person.
3. Ground truth label: urinary sugar level $y_i \in \{1,2,3,4,5, \text{NaN}\}$, where NaN indicates that the ground truth is missing. It is observed that 70% of the training data contain the target label (urinary sugar level).

Given the scenario above, please answer the following questions:

Q1.1: What do you think are the two biggest technical challenges to use machine learning for solving this problem? Please briefly describe them with a few sentences. For example, applicants can pick challenges from the following list below. It is also allowed to choose other challenges that are not listed here.

- Missing features
- Data size difference
- Domain mismatch
- Label imbalance

Q1.2: Which type of missing feature is the closest one in this scenario and why do you think it is the case?

- Missing completely at random (MCAR)
- Missing at random (MAR)
- Missing not at random (MNAR) (aka., Not missing at random (NMAR))

More information about the definitions of MCAR, MAR, and MNAR can be found here:

- <https://stefvanbuuren.name/fimd/sec-MCAR.html>
- <https://www.ncbi.nlm.nih.gov/books/NBK493614/>

Question 2

Should data preprocessing by imputing missing features always improve the performance for supervised classification? Please express your thoughts with an example.

Question 3

Describe the advantages and disadvantages of the following preprocessing methods by feature imputation for the supervised learning task (e.g., classification, regression):

- Discarding data points (i.e., a row in a DataFrame of pandas) that contains missing features and use only data points that have all features available.
- Discarding features (i.e., columns) that have missing values.
- Imputing the missing features with mode values of each categorical feature.
- Imputing the missing features with median values for each real-valued feature.
- Imputing the missing features with mean values of each real-valued feature.

Question 4

Discuss the weaknesses of tree-based methods (e.g., gradient boosting, random forest, decision tree) in tabular data and how deep learning could be promising to alleviate such weaknesses.

JE10. Research on machine learning methods for medical imaging

Choose one of the following three assignments and submit it as a pdf of 1~2 A4 pages. Please use a font of 10pt or larger. Note, medical images indicate images that are used in clinical settings such as X-ray, computed tomography (CT), magnetic resonance (MR) and pathological images.

- In clinical settings, It is useful to quantify the confidence of predictions made by deep learning based models. However, models can easily become overconfident if we do not take any measures, and methods such as confidence calibration have been in research to prevent such problems. Select one paper that proposes a method for calculating prediction confidence / uncertainty and summarize the content of the paper, give your opinion of the strength of the method and explain its disadvantages. You may refer to the following review paper if needed. Abdar et al, "A Review of Uncertainty Quantification in Deep Learning: Techniques, Applications and Challenges", <https://arxiv.org/abs/2011.06225>
- Recently, methods utilizing image synthesis which lead to efficient data collection have been proposed for general images (e.g. DatasetGAN*). In medical image analysis, it is generally difficult to collect large amounts of annotated data due to its expensiveness, and methods leading to efficient data collection have been in research. Select one paper which utilizes an image synthesis method which can potentially be used for medical image analysis (other than the DatasetGAN paper), and summarize the content of the paper, give your opinion of the strength of the method and explain its disadvantages. *Zhang et al, "DatasetGAN: Efficient Labeled Data Factory with Minimal Human Effort", <https://arxiv.org/abs/2104.06490>
- In medical image analysis, models tend to be less accurate with images that have been taken in different institutions / imaging conditions. In particular, differences in staining protocols can be problematic when constructing a model for pathological images that can generalize well. Select one paper that proposes a method for solving such domain adaptation problems and summarize the content of the paper, give your opinion of the strength of the method and explain its disadvantages. You may refer to the following review paper if needed. Guan et al, "Domain Adaptation for Medical Image Analysis: A Survey", <https://arxiv.org/abs/2102.09508>

JE11. Image processing and superresolution for remote sensing data

Pick a recent paper published after 2019 about image analysis, denoising or superresolution and review the paper from the following points. Preferably the target image should be Synthetic Aperture Radar(SAR) image, but is not restricted to it.

- i) Background of the study
- ii) What problem this study solves
- iii) How this study solves the problem
- iv) Strength of the study
- v) Weakness of the study

JE12. Depth estimation from video

Pick a recent paper published after 2019 about depth estimation by self-supervised learning and review the paper from the following points.

- i) Background of the study
- ii) What problem this study solves
- iii) How this study solves the problem
- iv) Strength of the study
- v) Weakness of the study

JE13. Machine learning of weather forecasting model based on atmospheric physics

Review the paper "Skilful precipitation nowcasting using deep generative models of radar" published from Nature 2021 (<https://www.nature.com/articles/s41586-021-03854-z>) from the following points.

- i) Background of the study
- ii) What problem this study solves
- iii) How this study solves the problem
- iv) Strength of the study
- v) Weakness of the study

JE14. Sensory Data Compression with Deep Learning

Choose one paper from the list below, and report the chosen paper. The report must include the following items.

- i) Background: why this study is conducted or required
- ii) Problem(s) that this paper solves
- iii) How the authors solve that problem(s)
- iv) main strengths of this paper

v) main weakness of this paper

vi) Suppose you are requested to apply this paper to the other sensors/data domains that are NOT examined in this paper. Argue what kind of sensors / data domain suit (or do not suit) to this paper, and why do you reason so?

Language: En/Ja

Page limit: A4 papers, 2-3 pages

Paper list:

- Zamir+, "Restormer: Efficient Transformer for Hig-resolution Image Restoration", CVPR 2022. <https://arxiv.org/abs/2111.09881>
- Hu+, "FVC: A New Framework towards Deep Video Compression in Feature space", CVPR 2021. <https://arxiv.org/abs/2105.09600v2>
- Cheng+, "Learned Image Compression with Discretized Gaussian Mixture Likelihoods and Attention Modules", CVPR 2020. https://openaccess.thecvf.com/content_CVPR_2020/html/Cheng_Learned_Image_Compression_With_Discretized_Gaussian_Mixture_Likelihoods_and_Attention_CVPR_2020_paper.html
- Chen+, "Simple Baselines for Image Restoration", <https://arxiv.org/abs/2204.04676>
- Cohen and Welling, "Steerable CNNs", ICLR 2017. <https://arxiv.org/pdf/1612.08498.pdf>
- Ulyanov+, "Deep Image Prior", CVPR 2018. https://dmitryulyanov.github.io/deep_image_prior

JE15. Interactive System for HCI for ML

Task 1

Please outline the research and development projects you have worked on in the following areas (250 - 500 words, 1 or 2 figures)

areas: Machine Learning, Computer Vision, Human-Computer Interaction, Accessibility, Visualization, Information Retrieval

Task 2

Please select and answer one from the following 1. 2. (400-800 words)

1. please select a paper about HCI for ML research from the following conferences and answer Q1-Q5 below.

Conference: CHI2019-2021, UIST2019-2021, IUI2019-2022

Q1 Why HCI for ML is necessary for the problem this research addresses?

Q2 What problem does the paper solve?

Q3 How does the paper solve the problem, and what are its features?

Q4 In what cases does the method work well, and in what cases does it not work well?

Q5 What are the inadequacies of this paper?

2. please answer the following Q1-3

Q1 How do you think HCI can contribute to machine learning problems? Please give at least three examples.

Q2 Please explain some ways to deal with problems about learning from small amounts of data.

Q3 Please explain some ways to make the acquisition and annotation of training data effective.

JE16. Just-in-Time Data Communication Protocol Development in a Heterogeneous Network Environment

Pick and read one from the paper list below, and answer the following questions. Note that your choice of paper will not affect the evaluation score.

- State the challenges to Quality of Service (QoS) and Quality of Experience (QoE) improvement by referring to the authors' description in the paper. Your answer may not be limited to those described in the paper.
- Describe the method(s) proposed in the paper and its/their characteristics and significances.
- Describe future work and/or potential problems of the proposed method(s) in deploying them onto the Internet with references to other papers or work. They must be known or pointed out in the other papers or work you refer to.

Paper list:

[1] M. Palmer et al., "VOXEL: Cross-layer Optimization for Video Streaming with Imperfect Transmission," ACM CoNEXT'21, pp.359--374, 2021

[2] Z. Zheng et al., "XLINK: QoE-driven multi-path QUIC transport in large-scale video services," ACM SIGCOMM'21, pp.418--432, 2021

JE17. Efficient Data Transfer Technology for Large-Scale Deep Learning

Pick and read one from the paper list below, and answer the following questions. Note that your choice of paper will not affect the evaluation score.

- State the challenges in distributed networking system, disaggregated computing, and data center networking, by referring to the authors' description in the paper. Your answer may not be limited to those described in the paper.
- Describe the method(s) proposed in the paper and its/their characteristics and significances.
- Describe future work and/or potential problems of the proposed method(s) for real-world deployment with references to other papers or work. They must be known or pointed out in the other papers or work you refer to.

Paper list:

[1] J. Min et al., "Gimbal: Enabling Multi-tenant Storage Disaggregation on SmartNIC JBOFs," ACM SIGCOMM'21, pp.106--122, 2021

[2] B. Li et al., "1Pipe: Scalable Total Order Communication in Data Center Networks," ACM SIGCOMM'21, pp.78--92, 2021

[3] S. Abdous et al., "Burst-tolerant datacenter networks with Vertigo," ACM CoNEXT'21, pp.1--15, 2021

[4] K.Liu et al., "Floodgate: taming incast in datacenter networks," ACM CoNEXT'21, pp.30--40, 2021

[5] R. Segal et al., "SOAR: minimizing network utilization with bounded in-network computing," ACM CoNEXT'21, pp.16--29, 2021

[6] Q. Zhang et al., "MimicNet: Fast Performance Estimates for Data Center Networks with Machine Learning," ACM SIGCOMM'21, pp.287--304, 2021

JE18. Application research on machine learning/molecular simulation for drug discovery

Please complete the following two assignments.

(1) Summarise an article.

Please choose and summarise one of the articles¹. You can either choose one of the following listed articles or pick any article related to the theme (ML/molecular simulation for drug design) by yourself². You might want to take into account the following points during summarization:

- The problem(s) the proposed method seeks to solve.
- Previous solutions and their shortcomings
- How the proposed method tries to solve the problem
- How the effectiveness of the proposed method is verified
- Limitations and possible drawbacks of the proposed method (either the authors' description or your own thoughts are OK)

(2) Describe your envision of a research area.

Pick one of the following research areas that you believe will advance in the next 10 years and describe your envision (perspective) of it.

- Molecular generative models
- Molecular property prediction
- Molecular simulation-based binding affinity prediction

You can focus on the drawbacks discussed in assignment (1) or on your own views. Please take into account the following two points:

- Reasons why the current technology has not addressed the limitations
- Some technical clues to address them.

This assignment is the starting point to draft the research topic you will work on during your internship. Therefore, please feel free to choose the topic/article you are particularly interested in.

Please report both assignments in one PDF file of 1 to 2 pages (A4) and use 10pt+ font for it.

Note:

1. You can assume readers have basic knowledge of these fields
2. Your choice of article will not affect your evaluation / selection.

Article list:

- Iterative Refinement Graph Neural Network for Antibody Sequence-Structure Co-design. <https://arxiv.org/abs/2110.04624>
- Benchmarking Peptide-Protein Docking and Interaction Prediction with AlphaFold-Multimer. <https://www.biorxiv.org/content/10.1101/2021.11.16.468810v1>
- EquiBind: Geometric Deep Learning for Drug Binding Structure Prediction. <https://arxiv.org/abs/2202.05146>
- Independent SE(3)-Equivariant Models for End-to-End Rigid Protein Docking. <https://arxiv.org/abs/2111.07786>
- Equivariant Diffusion for Molecule Generation in 3D. <https://arxiv.org/abs/2203.17003>
- Sample-Efficient Optimization in the Latent Space of Deep Generative Models via Weighted Retraining. <https://arxiv.org/abs/2006.09191>
- SpookyNet: Learning force fields with electronic degrees of freedom and nonlocal effects: <https://www.nature.com/articles/s41467-021-27504-0>

JE19. Applied research and development of machine learning and atomic simulation on materials

Please work on the following report assignment.

Identify one paper on a topic related to both atomic simulation and machine learning and summarize it from the following perspectives You may assume that the reader has a basic knowledge of the field.

- Bibliography of the paper
- The superiority of the paper over one or more of the previous studies cited in the paper
- The reason why the paper has its superiority

In your report, please report in PDF format and one page of A4 paper, using a font of 10pt or higher. For the selection of the paper, if you have an issue that you intend to work on in this internship, please use the relevant paper. If you do not yet have a firm topic in mind, you may choose a paper from the following list. Your selection of a particular paper will not make your evaluation any better or worse.

If you find it easier to work on the task of "JE06. Development of algorithms to calculate physical properties for Matlantis" rather than this task, you may work on that one instead.

- D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, Ab Initio Solution of the Many-Electron Schrödinger Equation with Deep Neural Networks, Phys. Rev. Research 2, 033429 (2020). <https://journals.aps.org/prresearch/abstract/10.1103/PhysRevResearch.2.033429>
- A. Chandrasekaran, D. Kamal, R. Batra, C. Kim, L. Chen, and R. Ramprasad, Solving the Electronic Structure Problem with Machine Learning, Npj Computational Materials 5, 22 (2019). <https://www.nature.com/articles/s41524-019-0162-7>
- O. T. Unke and M. Meuwly, PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges, Chem. Theory Comput., 15, 6, 3678–3693 (2019). <https://pubs.acs.org/doi/10.1021/acs.jctc.9b00181>
- S. Nikolov, M. A. Wood, A. Cangi, J.-B. Maillet, M.-C. Marinica, A. P. Thompson, M. P. Desjarlais, and J. Tranchida, Data-Driven Magneto-Elastic Predictions with Scalable Classical Spin-Lattice Dynamics, Npj Computational Materials 7, 1 (2021). <https://www.nature.com/articles/s41524-021-00617-2>

- B. Herzog, M. C. da Silva, B. Casier, M. Badawi, F. Pascale, T. Bučko, S. Lebègue, and Dario Rocca, Assessing the Accuracy of Machine Learning Thermodynamic Perturbation Theory: Density Functional Theory and Beyond, J. Chem. Theory Comput. 18 (3), 1382-1394 (2022).
<https://pubs.acs.org/doi/10.1021/acs.jctc.1c01034>
- M. Cools-Ceuppens, J. Dambre, and T. Verstraelen, Modeling Electronic Response Properties with an Explicit-Electron Machine Learning Potential, J. Chem. Theory Comput. 18, 1672 (2022).
<https://pubs.acs.org/doi/abs/10.1021/acs.jctc.1c00978>

JE20. Application of Deep Learning Techniques to Creation

(This theme's task is not open to the public.)

JE21. Deep learning for financial applications

General introduction

Choose one of the following three problems (A, B, and C) to answer. (You may answer more than one problem if you are interested, though it is not required for the selection process.)

- The report should be one or two pages of A4 sheets.
- Cite references appropriately.
- If necessary, please use any programming languages and/or existing software libraries. You do not have to submit source codes created for the assignment, but you may be asked to explain the details of the codes during the interview.

Problem A

A-1

What are the following three portfolio construction methods? Describe briefly (up to 5 lines for each). Please include "expected return(s)" and/or "covariance matrix" in the description.

- Mean-Variance portfolio
- Minimum variance portfolio
- Risk parity portfolio

A-2

Suppose you plan to invest in three risky assets, whose expected return vector μ and covariance matrix cov are estimated as follows:

```
mu = [0.005, 0.007, 0.004]

cov = [
    [1.0, 0.1, 0.2],
    [0.1, 2.0, 0.6],
```

```
[0.2, 0.6, 1.0]  
]
```

Create a program(s) to calculate two or more of the three portfolios given in A-1. Explain the implementation details (e.g., mathematical formulations and algorithms) using equations, etc. Calculate the portfolios using the above estimates of μ and cov , and discuss the results obtained.

Problem B

B-1

It is empirically well known that the volatilities of financial time series such as stock returns vary with time. Give an example of a multivariate time-series model that has time-varying volatility and correlation matrix, and describe its mathematical formulation.

B-2

Create a program that generates artificial data of multivariate time series. Generate a trivariate time-series data with length 1000 (which results in an array with size (1000, 3)), and describe the result with appropriate figures. Here, the generated data should have time-varying volatility and correlation matrix 'in some sense'. Also, explain what algorithm you used to generate the data. (NB: The algorithm you implement here does not have to be the same as the model you answered in B-1.)

Problem C

C-1

Describe briefly the following terminologies. Use equations if needed.

- Implied volatilities
- Volatility smiles

C-2

Create a program that calculates (European) call option prices under the Black-Scholes model. Calculate call option prices under the following setting, and describe the results.

- The current price of the underlying is 26750 JPY.
- The term to maturity is one month (= 1/12 year).
- Calculate option price for each strike value K in [26000, 26500, 27000, 27500, 28000] JPY.
- The annual volatility of the underlying is 20%.
- For simplicity, the annual risk-free rate is assumed to be 0%.

C-3

The following table shows the actual market prices C of the call options given in C-2.

Strike K	Option price C
------------	------------------

Strike K	Option price C
26000	1120
26500	760
27000	480
27500	270
28000	130

Create a program that calculates implied volatilities (under the Black-Scholes model). Compute the volatility smile for the market price data in the above, and discuss the result obtained.