

# Preferred Networks Internship 2020 Screening Task

---

There are two kinds of tasks: coding task and survey task. Please prepare and submit your solutions for both tasks.

NOTE: If your first choice is chip-related theme (i.e., either 25. Next-generation Architectures for Highly Efficient Learning and Inference or 26. Machine Learning and EDA), then please follow the instructions written in `README-chip-en.pdf`.

## Notice

Please do the task yourself. Do not share or discuss this coding 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 this 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.

## Things to submit

Please submit the following.

- Solution of coding task
  - Source code: Please create a directory called `src`. Put your code in this directory.
  - Report: Please name the file `report.pdf`.
  - Output for test cases: Please name the files `q1_out.txt` - `q3_out.txt`.
- Solution of survey task
  - Please name the file `survey.pdf`.

## How to submit

Create a zip archive with all of the submission materials and submit it on [this form](#). Due date is Friday, May 8th, 2019, 12:00 JST.

## First choice of themes

When solving the coding task and the survey task, please follow the first choice of themes when you applied.

If you want to change your first choice, please let us know your new first choice when you submit your solutions. In this case, please solve the coding and survey task based on the changed first choice theme.

## Inquiry

If you have any questions on this problem description, please contact us at [intern2020-admin@preferred.jp](mailto:intern2020-admin@preferred.jp). An updated version will be shared with all applicants if any change is made. Note that we cannot comment on the approach or give hints to the solution.

## Changelog

- Apr 24th, 2020 : Initial version

# Coding Task

---

In this task, we question basic coding ability.

Please create programs that solve the following questions Q1-Q3. Also, please verify that your program is working correctly by some means and summarize its result in a report.

During an evaluation, we will check not only the correctness of the submitted output but also the process of verifying the correctness of your program and the ability to verbalize the ideas.

## Program

Please use the programming language specified in the first choice theme (see next page).

Use only the standard libraries of each language for the purpose of testing your fundamental coding ability. You may use NumPy only for Python; libraries other than NumPy are not allowed.

You may put the source code in a separate file for each question, or you may put them all in one file for Q1-Q3.

We expect the runtime of the program is sufficiently short. In Q1-Q3, we assume that it takes less than a few seconds to process a single test case even when using an interpreted language such as Python.

We expect that the program is readable enough for engineers and researchers to understand the intent of the program. We don't necessarily require that your coding style is in line with any particular coding style.

## Verification

Verify that the program you have written works correctly. You may need to add new test cases or create a new program.

When you write a new program, please use the same language as the one you have solved in the task. In verification, you may use any libraries you like.

As a concrete situation, imagine a situation where your colleague asks you, for example, "Does the program you wrote really work?" We expect you to prepare material that you can answer to such a question with sufficient certainty.

## Report

Please submit a report containing the following information in a PDF format of no more than two A4 pages.

- How to run the program
- Explanation of how to solve the problem
  - Make sure to write an explanation that an engineer or researcher who doesn't know how to solve the problem can read and understand.
- Verification of the program
  - Summarize the above verification process.

## Evaluation criteria

Submissions will be evaluated on the following criteria.

- Readability: Program is sufficiently readable to understand its intention, and the solution is reasonably explained in report.
- Validity of verification: The correctness of the program is sufficiently verified.
- Correctness: The submitted output is correct, and the execution of the program is sufficiently quick.

## Programming language

Theme	Available programming language
1. Research and Development on Data Science / Machine Learning Application	Python3
2. Development of a Framework or Libraries for Executing Deep Learning Models on Actual Products	C / C++ / Rust
3. Development and Application of Differentiable Graphics and Rendering	For subtheme 1: Python3 / C++ For subtheme 2: Python3
4. Research and Development on Image Recognition	Python3
5. Development of Object Recognition System Using 3D Data	Python3
6. High-Performance Data Communication Network for Parallel and Distributed Deep Learning	Python3 / C++
7. Development of Compiler Backend for MN-Core	C++
8. Development of System Software for MN-Core	C++
9. Development of a Library for Multiple-types of Biological Data Analysis	Python3 / C++
10. Application Research or Development of Machine Learning or Molecular Simulation in Drug Discovery	Python3 / C++
11. Deep Learning-based Atomic simulation and Its Application to Material Development	Python3 / C++
12. Application of deep learning techniques to animation	Python3
13. Development of Simulator Developing Technologies and Simulator-related Technologies for Industrial Process	Python3 / C++ / Rust / Go
14. Application of Reinforcement Learning to Robotics	Python3
15. Computer Vision for Autonomous Driving	Python3
16. Development of Optuna	Python3
17. Development of Distributed Reinforcement Learning Technology	Python3 / C / C++
18. Development of Related Localization Technology	C++
20. Quantitative Finance Using Machine Learning	Python3 / C++ / Rust
21. Machine Learning-based Physical Simulation	Python3 / C++
22. Research on the Current Medical Image Problems Such As Practical Use of Few Samples and/or Evaluation of the Difference between Facilities/Machines	Python3
23. Machine Learning Research: Deep Learning towards Society	Python3
24. Quantitative Finance Using Machine Learning	Python3 / C++ / Rust

# Q1

---

You are given three integers  $x$ ,  $y$ , and  $d$ . Suppose that we will create a  $3 \times 3$  matrix where each element is either  $x$  or  $y$ , and its determinant is exactly  $d$ . Please compute the number of such matrices.

## Constraints

- $-10 \leq x \leq 10$
- $-10 \leq y \leq 10$
- $x \neq y$
- $-1000 \leq d \leq 1000$

## Input and output format

The input file is `q1_in.txt`. This file contains multiple test cases, with one line describing one test case. Each test case has integers  $x$ ,  $y$ , and  $d$  in this order, separated by a space. For each test case, output the answer in one line. The output file should be named `q1_out.txt`.

## Example input

```
0 1 2
0 1 99
```

## Output for example input

```
3
0
```

## Q2

---

We define strings  $S_1, S_2, \dots$  as follows.

- $S_1 = "a"$
- $S_2 = "b"$
- $S_3 = "c"$
- $S_k = S_{k-3} + S_{k-2} + S_{k-1}$  ( $k \geq 4$ ) (Here,  $+$  represents string concatenation.)

For example,  $S_4 = "abc"$ ,  $S_5 = "bcabc"$ .

You are given three integers  $k, p$ , and  $q$ . Please count the number of characters "a", "b", and "c" in the substring from the  $p$ -th to the  $q$ -th characters of  $S$ , respectively.

### Constraints

- $1 \leq k \leq 50$
- $1 \leq p \leq q \leq (\text{the length of } S_k)$

### Input and output format

The input file is `q2_in.txt`. This file contains multiple test cases, with one line describing one test case. Each test case has integers  $k, p$ , and  $q$  in this order, separated by a space. For each test case, output the answers in the format `a:(number of "a"),b:(number of "b"),c:(number of "c")` on one line. The output file should be named `q2_out.txt`.

### Example input

```
5 2 3
```

### Output for example input

```
a:1,b:0,c:1
```

## Q3

You are given two integers  $n$  and  $a_1$ . Now,  $n$  chairs are aligned in a straight line at equal intervals from left to right. Here are  $n$  people coming in turn and trying to sit in these chairs with the following rules.

- The first person sits in the chair  $a_1$  from the left.
- The second or subsequent person will sit in the chair farthest away from the one that is already occupied. If there is more than one such chair, the person sits in the chair at the far left of them.

Please compute which chair each person will sit in. Since the number of chairs in this problem is large, please calculate the following values:

Suppose that the  $i$ -th person will sit in the chair  $a_i$  from the left. ( $i=1,2,\dots,n$ ) Output the sum of even terms ( $a_2 + a_4 + a_6 + \dots$ ).

### Constraints

- $2 \leq n \leq 1,000,000$
- $1 \leq a_1 \leq n$

If you can't think of a computationally efficient solution for the above constraints, write and submit a program assuming that  $n \leq 100$ . You can ignore the input of  $n > 100$  in `q3_in.txt`.

### Input and output format

The input file is `q3_in.txt`. This file contains multiple test cases, with one line describing one test case. Each test case has integers  $n, a_1$  in this order, separated by a space. For each test case, output the answer in one line. The output file should be named `q3_out.txt`.

### Example input

```
6 2
```

### Output for example input

```
12
```

In this case,  $(a_1, a_2, a_3, a_4, a_5, a_6) = (2, 6, 4, 1, 3, 5)$ . Summing up even terms, we have  $6+1+5=12$ .

# Survey 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 in an A4 PDF format.

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, and the format is assumed to be free.

## 1. Research and Development on Data Science / Machine Learning Application

---

Describe the chance and challenges of machine learning technology in practical/social problem-solving. The essay should consist of the following parts:

- Describe one of the actual/social issues you think that machine learning technology is not yet fully utilized;
- Describe a machine learning-based solution for the problem and explain its significance;
- Describe the expected difficulties and issues in practice.

Submission guideline: Cite related work appropriately. You may include figures/tables if necessary. The paper size is A4, up to 2 pages, including tables, figures, and references.

## 2. Development of a Framework or Libraries for Executing Deep Learning Models on Actual Products

---

Please solve one of the following tasks based on sub-theme you want to choose.

- (for theme1) Explain steps for importing an ONNX model, converting it to a binary, and running the binary on an edge device, using the below vendor tools.
- Target model: <https://github.com/onnx/models/tree/master/vision/classification/mnist>
- Target vendor tools: choose one among CoreML, TensorRT, TFLite, Torch Mobile
- Following Keywords: Optimization(quantization, pruning, kernel fusion, etc.), Which device(CPU, GPU, DSP, etc.), Latency, ...
- (for theme2) Explain implementation of TLS(thread-local storage), especially overview about keyword "TLSDESC".

## 3. Development and Application of Differentiable Graphics and Rendering

---

Please solve one of the following tasks based on sub-theme you want to choose.

- (For theme 1) Please summarize the characteristics, the current development status, and the possible improvements you think of PyTorch3D, Kaolin, and TensorFlow Graphics.
- (For theme 2) Choose one paper which is related to the internship themes ('3D Reconstruction of Large Scene Using Differentiable Rendering Engine'), and published in CVPR2019, ICCV2019, or SIGGRAPH2019 main conference. Explain it in about one page (A4 size, free format, maximum of two pages) in Japanese or English. You can assume that readers have basic knowledge of computer vision.  
In particular, discuss the following things:
  - The problem the paper wants to solve
  - Existing solutions and drawbacks (why they cannot solve the problem)
  - The solution of the paper
  - How the paper verifies the effectiveness of the solution.
  - Limitations and potential drawbacks of the solution.

## 4. Research and Development on Image Recognition

---

Choose one paper which is related to the internship themes ('NAS for image recognition', or 'Image recognition with minimal annotation'), and published in CVPR2019 or ICCV2019 main conference. Explain it in about one page (A4 size, free format, maximum of two pages) in Japanese or English. You can assume that readers have basic knowledge of computer vision.

In particular, discuss the following things:

- The problem the paper wants to solve
- Existing solutions and drawbacks (why they cannot solve the problem)
- The solution of the paper
- How the paper verifies the effectiveness of the solution.
- Limitations and potential drawbacks of the solution.

## 5. Development of Object Recognition System Using 3D Data

---

Choose one paper which is related to the internship themes ('Object Surface Material Estimation on 3D Scanner', 'Training Image Recognition Model Efficiently by Using CG Images', or 'Developing Annotation Tools for 3D Data'), and published in CVPR2019, ICCV2019, or SIGGRAPH2019 main conference. Explain it in about one page (A4 size, free format, maximum of two pages) in Japanese or English. You can assume that readers have basic knowledge of computer vision.

In particular, discuss the following things:

- The problem the paper wants to solve
- Existing solutions and drawbacks (why they cannot solve the problem)
- The solution of the paper
- How the paper verifies the effectiveness of the solution.
- Limitations and potential drawbacks of the solution.

## 6. High-Performance Data Communication Network for Parallel and Distributed Deep Learning

---

1. Explain a static and dynamic routing using the following keywords. (path vector, distance vector, link state, ECMP)
  2. Explain the need for dynamic routing in data center networks and the problems it solves, both in terms of topology and protocol using RIFT (<https://datatracker.ietf.org/wg/rift/about/>, <https://datatracker.ietf.org/meeting/100/materials/slides-100-dcrouting-4-rift-routing-in-fat-trees/>) as an example.
  3. Point out issues RIFT has not solved (Optional)
- The entire document should be no more than two A4 pages

## 7. Development of Compiler Backend for MN-Core

---

Answer both 1. and 2.

1. Explain reasons with specific optimization methods: why automatic parallelization and compiling tools, adopted own DSL such as Halide, can provide more advanced optimization than general programming compilers.
2. Halide and oneDNN(ex-DNNL(ex-MKL-DNN)) adopt JIT compilation for optimization. Explain why static compilation is not enough.

## 8. Development of System Software for MN-Core

---

Answer the following questions about user space IO (UIO).

1. its advantages and disadvantages, and what areas to good for use
2. give specific recent research that uses it and explain its contribution.

## 9. Development of a Library for Multiple-types of Biological Data Analysis

---



Choose one paper from the following journals or conferences published within three years and explain it in about one page (A4 size, free format, maximum of two pages) in Japanese or English. You can assume that readers have basic knowledge of bioinformatics (undergraduate level of department of Biology and/or Informatics)

Journal or conferences

- Bioinformatics, ISMB/ECCB, RECOMB  
In particular, discuss the following things:
- The problem the paper wants to solve
- Existing solutions and drawbacks (why they cannot solve the problem)
- The solution of the paper
- How the paper verifies the effectiveness of the solution.
- Limitations and potential drawbacks of the solution.

## 10. Application Research or Development of Machine Learning or Molecular Simulation in Drug Discovery

---

1. Please choose one paper that combines the deep learning and biological molecule simulation, including: Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning, Noé et al., Science 365, eaaw1147 (2019)  
<http://dx.doi.org/10.1126/science.aaw1147>  
<https://arxiv.org/abs/1812.01729>
  2. Summarize the excel points over the previous works and state the position of the paper in the research field.
  3. Point possible flaws of this paper, and if possible, suggest some ways to improve the quality of the paper.
- We expect the report is within two A4 pages. Any format is acceptable.

## 11. Deep Learning-based Atomic simulation and Its Application to Material Development

---

1. Choose 2 papers, which attempt to apply machine learning into the atomistic simulation, published on or after 2017. Explain those 2 papers by comparing them. Concretely, summarize novel suggestions and pros/cons of each paper objectively.
2. The simulation of the atomic system using machine learning is getting attention recently. Suggest an issue which you think it will be solved within 5 years. For example, focusing on the disadvantages of the paper in task 1, or it is also applicable to write your own opinion about new material search. Please summarize the current issues with the reason why it is not achieved and describe the hint to solve the issue.

Your essay must not exceed two A4-sized pages in total (format-free).

## 12. Application of deep learning techniques to animation

---

Please choose a paper related to the application of deep learning techniques to animation and write an essay that covers:

1. A brief summary of the paper
2. Why you find this paper exciting, based on its technical contributions, novelty, application, etc.
3. If you are going to either publish an impactful paper or build a practical application during the internship program, how will you utilize the ideas in this paper, what are some potential difficulties, and how will you overcome these problems?

Your essay must not exceed two A4-sized pages (format-free).

We strongly recommend you to choose the paper from the following lists:

- CVPR 2019 <http://openaccess.thecvf.com/CVPR2019.py>
- ICCV 2019 <http://openaccess.thecvf.com/ICCV2019.py>
- CVPR 2020 (the list is not available yet, but some authors have published their works on the Internet)

You may also choose a paper that is not listed above. In that case, please also justify your choice and explain how you will utilize it to solve practical problems related to animation.

## 13. Development of Simulator Developing Technologies and Simulator-related Technologies for Industrial Process

---

In recent years, researchers have proposed differentiable programming frameworks such as Jax [1] and DiffTaichi [2]. Consider the following questions about differentiable programming and its design for physical simulation.

Question 1. Explain the paper of DiffTaichi [2], a relatively new proposal: Summarize the paper in your words while revealing the features of differentiable programming; And criticize the work from the viewpoint of using it for physical simulation, clarifying the advantages and disadvantages.

Question 2. Describe your thoughts on the functions and design that a differentiable programming framework should have to run a physical simulation. Here is a list of topic examples: you may discuss one of the following topics or one is not listed here.

- Language design: language construct for automatic derivation, language construct for parallelization, type system to ensure programs to be physically meaningful, prevention of perturbation confusion, etc.
- Compiler design: required path, intermediate representation, etc.
- Runtime design: management of data structure for automatic derivation, the realization of parallelization, garbage collection, etc.

Submission guideline: Cite related work appropriately. You may include figures/tables if necessary. The paper size is A4, up to 2 pages (one page for Q1 and another for Q2), including tables, figures, and references.

[1] Roy Frostig, et al. Compiling machine learning programs via high-level tracing. SysML Conference 2018. <https://mlsys.org/Conferences/2019/doc/2018/146.pdf>

[2] Yuanming Hu, et al. DiffTaichi: Differentiable programming for physical simulation. arXiv preprint, 2019. <https://arxiv.org/abs/1910.00935>

## 14. Application of Reinforcement Learning to Robotics

---

1. Choose a recent paper opened within two years, which is related to the theme, i.e., RL for robotics
2. Summarize the good points of this paper and state the position of the paper in the literature.
3. Point possible flaws of this paper and suggest some ways to improve the quality of the paper if possible.

We expect the report is within two A4 pages. Any format is OK.

## 15. Computer Vision for Autonomous Driving

---

1. Choose a recent paper opened within two years, which is related to the work you want to do during the internship, i.e., something related to the autonomous driving CV task such as segmentation, object detection, tracking, etc.
2. Summarize the good points of this paper and state the position of the paper in the literature.
3. Point possible flaws of this paper and suggest some ways to improve the quality of the paper if possible.

We expect the report is within two A4 pages. Any format is OK.

## 16. Development of Optuna

---

Do one of the following assignments.

[Assignment A]

In hyperparameter optimization, oftentimes, a small subset of hyperparameters is dominant in how they impact the performance of a model. Describe one known method for assessing hyperparameter importance and explain possible limitations, theoretical and/or practical, and how they could be addressed.

[Assignment B]

Describe the performance bottlenecks and/or any other problem with the [RDBStorage](#) in Optuna v1.3.0.

If possible, try to address these problems. You can extend the RDBStorage interface and change the database schema if needed.

## 17. Development of Distributed Reinforcement Learning Technology

---

Compare IMPALA [1] and SEED RL [2] and discuss the following two topics:

1. Differences between Distributed Learning of Common Image Classification and Distributed Reinforcement Learning
2. About these system architectures and algorithms

Also, describe the issues and improvements of these systems.

[1] Espeholt, L., et.al. (2018). IMPALA: Scalable Distributed Deep-RL with Importance Weighted Actor-Learner Architectures.

35th International Conference on Machine Learning, ICML 2018, 4, 2263–2284.

arxiv URL: <http://arxiv.org/abs/1802.01561>

[2] Espeholt, L., et.al. (2019). SEED RL: Scalable and Efficient Deep-RL with Accelerated Central Inference.

arxiv URL: <http://arxiv.org/abs/1910.06591>

## 18. Development of Related Localization Technology

---

1. Choose the theme you want to work related to Localization, and select one related paper (preferably a paper published within two years). If you choose the direct / indirect hybrid SLAM development, please select one of the following two papers.

Direct Sparse Odometry: <https://vision.in.tum.de/research/vslam/dso>

LSD-SLAM: Large-Scale Direct Monocular SLAM: <https://vision.in.tum.de/research/vslam/lslam>

2. Summarize the good points of this paper and state the position of the paper in the literature.
  3. Point possible flaws of this paper and suggest some ways to improve the quality of the paper if possible.
- We expect the report is within two A4 pages. Any format is OK.

## 20. Quantitative Finance Using Machine Learning

---

Please choose one of the following questions 1 and 2 and answer.

There is no difference between Question 1 and Question 2 from the viewpoint of evaluation.

In the report, please use PDF format and 1 or 2 pages of A4 paper using fonts of 10pt or more. In addition to the report, you can attach your programs in PDF format as an appendix after the report.

### Question 1

The following Prof. Kenneth R. French website distributes financial return data called factors. Please select one or more of these factors and analyze and report the qualitative and quantitative features that will be useful when making a profit.

[http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html)

### Question 2

Please select one of the following papers and explain it on one page of A4 paper. The papers are listed in alphabetical order by the first author's last name. The choice of the paper does not affect your evaluation, so choose the paper that is best for you. In your explanation, please give more emphasis on practical application rather than theoretical aspects.

[1] Jushan Bai and Serena Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, January 2002.

[2] Fischer Black and Robert Litterman. Global portfolio optimization. *Financial Analysts Journal*, 48(5):28–43, September 1992.

[3] Alan Brace, Dariusz Gatarek, and Marek Musiela. The market model of interest rate dynamics. *Mathematical Finance*, 7(2):127–155, April 1997.

[4] Lawrence J. Christiano, Martin Eichenbaum, and Charles L. Evans. Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1):1–45, February 2005.

[5] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6):933–949, March 2018.

[6] Marc Hoffmann, Mathieu Rosenbaum, and Nakahiro Yoshida. Estimation of the lead-lag parameter from non-synchronous data. *Bernoulli*, 19(2):426–461, May 2013.

[7] Lutz Kilian. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99(3):1053–1069, May 2009.

- [8] Hyunjik Kim, Andriy Mnih, Jonathan Schwarz, Marta Garnelo, Ali Eslami, Dan Rosenbaum, Oriol Vinyals, and Yee Whye Teh. Attentive neural processes. In International Conference on Learning Representations, 2019.
- [9] Nobuhiro Kiyotaki and John Moore. Credit cycles. *Journal of Political Economy*, 105(2):211–248, April 1997.
- [10] Hayne E. Leland. Option pricing and replication with transactions costs. *The Journal of Finance*, 40(5):1283–1301, December 1985.
- [11] Song Liu, Makoto Yamada, Nigel Collier, and Masashi Sugiyama. Change-point detection in timeseries data by relative density-ratio estimation. *Neural Networks*, 43:72–83, July 2013.
- [12] Ziyin Liu, Zhikang Wang, Paul Pu Liang, Russ R Salakhutdinov, Louis-Philippe Morency, and Masahito Ueda. Deep gamblers: Learning to abstain with portfolio theory. In *Advances in Neural Information Processing Systems 32*, pages 10623–10633. 2019.
- [13] Zichao Long, Yiping Lu, Xianzhong Ma, and Bin Dong. PDE-net: Learning PDEs from data. In *Proceedings of the 35th International Conference on Machine Learning*. PMLR, July 2018.
- [14] R. David Mclean and Jeffrey Pontiff. Does academic research destroy stock return predictability? *The Journal of Finance*, 71(1):5–32, January 2016.
- [15] Sorin Solomon and Peter Richmond. Power laws of wealth, market order volumes and market returns. *Physica A: Statistical Mechanics and its Applications*, 299(1-2):188–197, October 2001.
- [16] Jingyuan Wang, Yang Zhang, Ke Tang, Junjie Wu, and Zhang Xiong. Alphastock: A buying-winners-and-selling-lossers investment strategy using interpretable deep reinforcement attention networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019.
- [17] Your favorite paper published as of May 1, 2020

## 21. Machine Learning-based Physical Simulation

---

Describe your thoughts on how to reproduce and predict real-world physical phenomena on a computer. The essay should consist of the following parts:

1. Give a brief explanation of the physical events that you want to reproduce or predict on a computer;
2. Explain a method to simulate the physical events by numerical calculation based on scientific theory;
3. Explain one of the existing methods to predict the physical events by machine learning approach;
4. Describe your thoughts on the significance, challenges, and prospects of using machine learning technology to express, reproduce, or predict the physical events on a computer.

Submission guideline: Cite related work appropriately. You may include figures/tables if necessary. The paper size is A4, up to 2 pages (one page for parts 1-3 and another for 4), including tables, figures, and references.

## 22. Research on the Current Medical Image Problems Such As Practical Use of Few Samples and/or Evaluation of the Difference between Facilities/Machines

---

Due to the differences between facilities and machines, it is generally difficult to collect a lot of homogenous data on medical image analysis. Another difficulty is to collect a sufficient amount of annotated data due to the high cost of annotation that requires a time-consuming process by medical professionals.

Choose one or more technical papers to approach either/both these issues and write its (their) contributions and the room for improvements within 1 or 2 pages of A4 sheets.

Examples of the medical images include (but are not limited to) X-rays, CT/MRI scans and histological images.

## 23. Machine Learning Research: Deep Learning towards Society

---

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 2019 <http://proceedings.mlr.press/v97/>
- AISTATS 2019 <http://proceedings.mlr.press/v89/>

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., <http://proceedings.mlr.press/v97/abcdefg19.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).

## 24. Quantitative Finance Using Machine Learning

Please choose one of the following questions 1 and 2 and answer.

There is no difference between Question 1 and Question 2 from the viewpoint of evaluation.

In the report, please use PDF format and 1 or 2 pages of A4 paper using fonts of 10pt or more. In addition to the report, you can attach your programs in PDF format as an appendix after the report.

### Question 1

The following Prof. Kenneth R. French website distributes financial return data called factors. Please select one or more of these factors and analyze and report the qualitative and quantitative features that will be useful when making a profit.

[http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html)

### Question 2

Please select one of the following papers and explain it on one page of A4 paper. The papers are listed in alphabetical order by the first author's last name. The choice of the paper does not affect your evaluation, so choose the paper that is best for you. In your explanation, please give more emphasis on practical application rather than theoretical aspects.

- [1] Jushan Bai and Serena Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, January 2002.
- [2] Fischer Black and Robert Litterman. Global portfolio optimization. *Financial Analysts Journal*, 48(5):28–43, September 1992.
- [3] Alan Brace, Dariusz Gatarek, and Marek Musiela. The market model of interest rate dynamics. *Mathematical Finance*, 7(2):127–155, April 1997.
- [4] Lawrence J. Christiano, Martin Eichenbaum, and Charles L. Evans. Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1):1–45, February 2005.
- [5] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6):933–949, March 2018.
- [6] Marc Hoffmann, Mathieu Rosenbaum, and Nakahiro Yoshida. Estimation of the lead-lag parameter from non-synchronous data. *Bernoulli*, 19(2):426–461, May 2013.
- [7] Lutz Kilian. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99(3):1053–1069, May 2009.
- [8] Hyunjik Kim, Andriy Mnih, Jonathan Schwarz, Marta Garnelo, Ali Eslami, Dan Rosenbaum, Oriol Vinyals, and Yee Whye Teh. Attentive neural processes. In *International Conference on Learning Representations*, 2019.
- [9] Nobuhiro Kiyotaki and John Moore. Credit cycles. *Journal of Political Economy*, 105(2):211–248, April 1997.
- [10] Hayne E. Leland. Option pricing and replication with transactions costs. *The Journal of Finance*, 40(5):1283–1301, December 1985.
- [11] Song Liu, Makoto Yamada, Nigel Collier, and Masashi Sugiyama. Change-point detection in timeseries data by relative density-ratio estimation. *Neural Networks*, 43:72–83, July 2013.
- [12] Ziyin Liu, Zhikang Wang, Paul Pu Liang, Russ R Salakhutdinov, Louis-Philippe Morency, and

- Masahito Ueda. Deep gamblers: Learning to abstain with portfolio theory. In Advances in Neural Information Processing Systems 32, pages 10623–10633. 2019.
- [13] Zichao Long, Yiping Lu, Xianzhong Ma, and Bin Dong. PDE-net: Learning PDEs from data. In Proceedings of the 35th International Conference on Machine Learning. PMLR, July 2018.
- [14] R. David Mclean and Jeffrey Pontiff. Does academic research destroy stock return predictability? The Journal of Finance, 71(1):5–32, January 2016.
- [15] Sorin Solomon and Peter Richmond. Power laws of wealth, market order volumes and market returns. Physica A: Statistical Mechanics and its Applications, 299(1-2):188–197, October 2001.
- [16] Jingyuan Wang, Yang Zhang, Ke Tang, Junjie Wu, and Zhang Xiong. Alphastock: A buying-winners-and-selling-lossers investment strategy using interpretable deep reinforcement attention networks. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019.
- [17] Your favorite paper published as of May 1, 2020