

# Preferred Networks インターンシップ<sup>o</sup>2020

## 選考課題

コーディング課題と調査課題の2つがあります。両者について回答を作成し提出して下さい。

注: 第一希望テーマとしてチップ関係のテーマ( 25. 次世代の高効率学習・推論アーキテクチャー または 26. 機械学習とEDA)を選ぶ方はこの課題ではなく、別途添付の README-chip-ja.pdf に沿って課題を進めて下さい。

### 諸注意

課題には自分だけで取り組んでください。この課題を他の応募者を含めた他人と共有・相談することを禁止します。GitHub の公開リポジトリや SNS 等に解答や問題をアップロードする行為も禁止します。(選考期間終了後、インターンシップ運営が問題を公開する可能性があります。インターンシップ運営が問題を公開した後は、解答を公開したりしていただいても構いません。) 漏洩の証拠が見つかった場合、その応募者は失格となります。ある応募者が別の応募者に回答をコピーさせた場合、双方の応募者が失格となります。

想定所要時間はコーディング課題と調査課題の両方を合わせて最大2日です。全課題が解けていなくても提出は可能ですので、学業に支障の無い範囲で取り組んで下さい。

### 提出物

以下のものを提出してください。

- コーディング課題の解答
  - ソースコード: `src` という名前のディレクトリを作りその下にソースコードを置いてください
  - レポート: `coding.pdf` というファイル名で PDF にまとめてください
  - テストケースに対する出力結果: `q1_out.txt` - `q3_out.txt` というファイル名にしてください
- 調査課題の解答: `survey.pdf` というファイル名で PDF にまとめてください

### 提出方法

提出物を単一のパスワード無し zip ファイルにまとめ、[こちらの専用フォーム](#)より応募してください。締切は日本時間2020年5月8日(金) 12:00 (正午) です。

### 第一希望テーマについて

インターンシップ選考課題では応募時にお選びした第一希望テーマに沿って進めて頂くことになります。

もし何らかの理由で第一希望テーマを変更したい場合は、課題の提出時にフォームより新しい第一希望テーマをお知らせ下さい。変更した際には、変更後の第一希望テーマに基づいてコーディング課題と調査課題を解いて下さい。

### 問い合わせ

問題文に関して質問がある場合は [intern2020-admin@preferred.jp](mailto:intern2020-admin@preferred.jp) までご連絡ください。問題文に訂正が行われた場合は応募者全員にアナウンスいたします。なお、アプローチや解法に関する問い合わせにはお答えできません。

### 変更履歴

2020年4月24日: 初版

# コーディング課題

この課題では基礎的なコーディング能力を問います。

以下の問1-3を解くプログラムを作成してください。さらに、作成したプログラムが正しく動いていることを何らかの方法で検証し、それをレポートの形でまとめてください。

評価の際には単に提出された出力結果の正しさだけでなく、プログラムの正しさを確かめるプロセスや、考え方を言語化する能力も重視します。

## プログラム

コーディングに用いるプログラミング言語には、第一希望テーマの指定する言語を用いてください(次のページ参照)。

コーディング能力を見るという目的のため、各言語の標準ライブラリのみを使用して下さい。Python に関するのみ、NumPy を使用してもよいです。NumPy 以外のライブラリの使用は禁止とします。

ソースコードは問題ごとに別ファイルにしても構いませんし、問1-3ですべて1ファイルにまとめても構いません。

プログラムは十分高速であることを期待します。問1-3で、Python のようなインタプリタ型の言語を使った場合でも1つのテストケースを処理するのに数秒以内で終わることを想定します。

エンジニアやリサーチャーが読んでプログラムの意図を理解できる程度の可読性を目指してください。コーディングスタイルが特定のコーディング規約に沿っていることは必ずしも要求しません。

## 検証

あなたの書いたプログラムが正しく動作することを確認してください。これには、新たにテストケースを追加したり、新たにプログラムを作成する必要があるかもしれません。また、解法のプログラム自体も検証がしやすいように作っておく必要があるかもしれません。

新たにプログラムを作成する際、検証で用いる言語は課題で解いたものと同じものを用いて下さい。ライブラリは自由なものを使用していただいて構いません。

具体的な状況として、例えばあなたの同僚に「あなたの書いたプログラムは本当に正しく動作しているのか」などと問われたときに十分な確度をもって答えられる資料を作ることを想像してください。

## レポート

以下の内容を A4 2ページ以内の PDF にレポートとしてまとめ、提出してください。

- プログラムの実行方法
- 問題の解法説明
  - 問題の解法を知らないエンジニアやリサーチャーが読んでも理解できるレベルの説明を書くことを意識してください。
- プログラムの正しさの検証
  - 上記の検証過程をまとめてください。

## 評価基準

提出物は以下の基準で評価します。

- 可読性: プログラムに意図を理解できる程度の可読性があり、またレポート全体が妥当な根拠をもって説明されていること
- 検証の妥当性: プログラムの正しさについて十分検証できていること
- 正当性: 正しい解を出力できており、プログラムの実行速度が十分であること

## プログラミング言語

テーマ	利用可能なプログラミング言語
1. データサイエンス・機械学習応用に関する研究開発	Python3
2. 深層学習モデルを社会実装するためのフレームワーク、ライブラリ開発	C / C++ / Rust
3. 微分可能レンダラの開発と応用	サブテーマ1の場合: Python3 / C++ サブテーマ2の場合: Python3
4. 一般の画像認識のための技術開発	Python3
5. 三次元データを活用した物体認識技術開発	Python3
6. 分散機械学習基盤におけるデータ通信ネットワークの改善	Python3 / C++
7. MN-Core向けコンパイラバックエンドの開発	C++
8. MN-Core向けシステムソフトウェアの開発	C++
9. 種々の医療データ解析のためのライブラリ開発	Python3 / C++
10. 創薬に関する機械学習や分子シミュレーションの開発・応用研究	Python3 / C++
11. 深層学習に基づく原子系シミュレータと材料開発への応用	Python3 / C++
12. ディープラーニングを使ったアニメーションへの応用	Python3
13. 工業プロセスのシミュレータ実装技術および周辺技術の開発	Python3 / C++ / Rust / Go
14. 強化学習のロボティクスへの応用	Python3
15. 自動運転のためのコンピュータビジョン	Python3
16. Optuna 開発	Python3
17. 分散強化学習の技術開発	Python3 / C / C++
18. Localizationに関する技術開発	C++
20. 機械学習を用いた数量ファイナンス	Python3 / C++ / Rust
21. 機械学習ベースの物理シミュレーションモデル開発	Python3 / C++
22. 医用画像を用いた機械学習の種々の問題に関する研究	Python3
23. ディープラーニングの社会実装に向けた機械学習の研究	Python3
24. 機械学習を用いた数量ファイナンス	Python3 / C++ / Rust

# 問1

3つの整数  $x, y, d$  が与えられる。3x3 行列の各要素を  $x$  または  $y$  で埋めるとき、行列式がちょうど  $d$  になるような埋め方は何通りあるか数えよ。

## 制約

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

## 入出力形式

入力ファイルは `q1_in.txt` である。このファイルには複数のテストケースが含まれ、1行には1つのテストケースが記されている。各テストケースは整数  $x, y, d$  がこの順番でスペース区切りで並べられている。それぞれのテストケースに対して解答を1行で出力せよ。出力ファイル名は `q1_out.txt` である。

## サンプル入力

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

## サンプル入力に対する出力

```
3
0
```

## 問2

文字列  $S_1, S_2, \dots$  を以下で定める。

- $S_1 = "a"$
- $S_2 = "b"$
- $S_3 = "c"$
- $S_k = S_{k-3} + S_{k-2} + S_{k-1}$  ( $k \geq 4$ ) (ここで、 $+$  は文字列の結合を表す)

例えば、 $S_4 = "abc"$ ,  $S_5 = "bcabc"$  である。

3つの整数  $k, p, q$  が与えられる。 $S_k$  の  $p$  文字目から  $q$  文字目までに含まれる文字  $"a", "b", "c"$  の個数をそれぞれ求めよ。

### 制約

- $1 \leq k \leq 50$
- $1 \leq p \leq q \leq (S_k \text{ の長さ})$

### 入出力形式

入力ファイルは `q2_in.txt` である。このファイルには複数のテストケースが含まれ、1行には1つのテストケースが記されている。各テストケースは整数  $k, p, q$  がこの順番でスペース区切りで並べられている。それぞれのテストケースに対して解答を `a: ("a" の個数), b: ("b" の個数), c: ("c" の個数)` というフォーマットで1行で出力せよ。出力ファイル名は `q2_out.txt` である。

### サンプル入力

```
5 2 3
```

### サンプル入力に対する出力

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

## 問3

2つの整数  $n, a_1$  が与えられる。いま、 $n$  個の椅子が左右一直線上に等間隔で並べられている。ここに  $n$  人の人々が順番にやってきて以下のルールでこの椅子に座ろうとしている。

- 最初の人 は左から  $a_1$  番目の椅子に座る。
- 2番目以降の人はいま空いている席のうち、既に誰かが座っている席からもっとも離れた席に座る。そのような席が複数ある場合は、それらのうち最も左端にある椅子に座る。

それぞれの人がどの椅子に座ることになるか求めて欲しい。しかし全ての結果を出力すると出力ファイルが大きくなるため、代わりに以下の値を計算して出力せよ:

$i$  番目の人は左から  $a_i$  番目に座るものとする ( $i=1,2,\dots,n$ )。偶数項の和 ( $a_2 + a_4 + a_6 + \dots$ ) を出力せよ。

### 制約

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

もし上記の制約について計算効率の良い解法が思いつかない場合は、 $n \leq 100$  であるとしてプログラムを書いて提出してください。 `q3_in.txt` の中で  $n > 100$  である入力については無視して頂いて構いません。

### 入出力形式

入力ファイルは `q3_in.txt` である。このファイルには複数のテストケースが含まれ、1行には1つのテストケースが記されている。各テストケースは整数  $n, a_1$  がこの順番でスペース区切りで並べられている。それぞれのテストケースに対して解答を1行で出力せよ。出力ファイル名は `q3_out.txt` である。

### サンプル入力

```
6 2
```

### サンプル入力に対する出力

```
12
```

$(a_1, a_2, a_3, a_4, a_5, a_6) = (2, 6, 4, 1, 3, 5)$  となる。偶数項だけ足すと  $6+1+5=12$  となる。

# 調査課題

この課題ではインターンのテーマに沿った専門知識の理解度を問います。あなたが選んだ第一希望テーマに相当する設問を以下の中から選び、その解答を A4 サイズの PDF にまとめて提出して下さい。

ページ制限、様式については各設問で指定された形式に従って下さい。特に指定がない場合はページ数は2枚以内、様式は自由とします。

## 1. データサイエンス・機械学習応用に関する研究開発

機械学習技術を用いて社会的課題を解決することの可能性と課題について、つぎの要素を含めながら議論してください。

- ・機械学習技術がまだ十分に活用されていないとあなたが考える社会的課題を1つ取り上げ、説明してください。
- ・機械学習のアプローチによるその課題の解決策を1つ述べ、その解決策の意義を説明してください。
- ・さらにその適用・実施において想定される困難や課題について説明してください。

関連研究を適切に引用してください。必要ならば図表を用いても構いません。図表や参考文献を含め、A4用紙2ページ以内でレポートを作成してください。

## 2. 深層学習モデルを社会実装するためのフレームワーク、ライブラリ開発

希望するサブテーマに応じて以下のいずれかの課題を解いて下さい。

- ・(テーマ①) ONNXモデルを各ベンダツールを使用してデバイス向けバイナリに変換し実行するまでの流れをレポートしてください。
  - 対象のモデル: <https://github.com/onnx/models/tree/master/vision/classification/mnist>
  - 対象のベンダツール: CoreML, TensorRT, TFLite, Torch Mobile から1つ選択
  - レポートに含めるキーワード: 最適化(量子化、枝刈り, kernel fusionなど), どのデバイスか(CPU/GPU/DSPなど), 推論の実行速度, 左記以外について調べても構いません
- ・(テーマ②) TLS(thread local storage)がどう実現されているかについて調べ、特に TLSDESC というキーワードについて、概要をレポートしてください

## 3. 微分可能レンダラの開発と応用

希望するサブテーマに応じて以下のいずれかの課題を解いて下さい。

- ・テーマ1 PyTorch3D, Kaolin, TensorFlow Graphics の特徴と開発状況、考えられる改良についてまとめてください。
- ・テーマ2: CVPR2019, ICCV2019, SIGGRAPH2019の本会議で発表された論文のなかからインターンシップのテーマである「微分可能レンダラを用いた部屋全体の三次元再構成」に関連するものを1つ選び、その内容をA4 1枚程度（形式自由，最大A4 2枚）で日本語または英語で説明して下さい。読者は画像認識に関する基礎的な知識を持っていると仮定して構いません。  
特に以下の内容について論じてください
  - 論文が解決したい問題
  - 過去の解決方法とその欠点（なぜその問題を解決できないか）
  - 論文がその問題をどのように解決しているか
  - 解決方法の有効性をどのように検証しているか
  - 解決方法の限界と考えられる欠点

## 4. 一般の画像認識のための技術開発

CVPR2019, ICCV2019の本会議で発表された論文のなかからインターンシップのテーマである「画像認識へのNASの応用」または「省アノテーションコストでの画像認識」に関連するものを1つ選び、その内容をA4 1枚程度（形式自由，最大A4 2枚）で日本語または英語で説明して下さい。読者は画像認識に関する基礎的な知識を持っていると仮定して構いません。

特に以下の内容について論じてください

- 論文が解決したい問題
- 過去の解決方法とその欠点（なぜその問題を解決できないか）
- 論文がその問題をどのように解決しているか
- 解決方法の有効性をどのように検証しているか
- 解決方法の限界と考えられる欠点

## 5. 三次元データを活用した物体認識技術開発

CVPR2019, ICCV2019, SIGGRAPH2019の本会議で発表された論文のなかからインターンシップのテーマである「物体の3Dスキャナにおけるマテリアル推定技術開発」「3次元データに対するアノテーションツールの開発」または「CGデータを活用した画像認識器の学習」に関連するものを1つ選び、その内容をA4 1枚程度（形式自由、最大A4 2枚）で日本語または英語で説明して下さい。読者は画像認識に関する基礎的な知識を持っていると仮定して構いません。

特に以下の内容について論じてください

- 論文が解決したい問題
- 過去の解決方法とその欠点（なぜその問題を解決できないか）
- 論文がその問題をどのように解決しているか
- 解決方法の有効性をどのように検証しているか
- 解決方法の限界と考えられる欠点

## 6. 分散機械学習基盤におけるデータ通信ネットワークの改善

1. 静的経路制御、動的経路制御の仕組みについて以下キーワードを用いて説明せよ(path vector, distance vector, link state, ECMP)
2. データセンタネットワークにおけるdynamic routingについてRIFT(<https://datatracker.ietf.org/wg/rift/about/>, <http://datatracker.ietf.org/meeting/100/materials/slides-100-dcrouiting-4-rift-routing-in-fat-trees/>)を例にその必要性と解決している課題についてトポロジ/プロトコルの両面から説明せよ
3. RIFTでは解決できていないデータセンタネットワークに関する課題を挙げよ(Optional)  
全体でA4, 2枚以内で記述してください(様式自由)

## 7. MN-Core向けコンパイラバックエンドの開発

以下の1.と2.を回答してください。

1. 独自のDSLを定義するようなHalideなどの自動並列化コンパイラが、一般的な汎用言語コンパイラよりも高度な最適化を実現できる理由を、具体的な最適化の手法を挙げて説明せよ
2. HalideやoneDNN(旧DNNL(旧MKL-DNN))では最適化にJITコンパイルを採用しているが、なぜ静的なコンパイルでは不十分なのかを説明せよ

## 8. MN-Core向けシステムソフトウェアの開発

user space IOについて以下まとめよ

1. その利点と欠点、fitする領域がどういったものか
2. その活用について具体的に近年の研究を挙げその貢献を説明せよ

## 9. 種々の医療データ解析のためのライブラリ開発

以下の雑誌または会議で過去3年以内に出版された論文を1つ選び、その内容をA4 1枚程度（形式自由、最大A4 2枚）で日本語または英語で説明して下さい。読者はバイオインフォマティクスに関する基礎的な知識（生物学科・情報学科の学部程度）を持っていると仮定して構いません。

雑誌・会議候補

- Bioinformatics, ISMB/ECCB, RECOMB  
特に以下の内容について論じてください
- 論文が解決したい問題
- 過去の解決方法とその欠点（なぜその問題を解決できないか）
- 論文がその問題をどのように解決しているか



- 解決方法の有効性をどのように検証しているか
- 解決方法の限界と考えられる欠点

## 10. 創薬に関する機械学習や分子シミュレーションの開発・応用研究

### 1. 以下の論文

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>

あるいは、深層学習と生体分子のシミュレーションを組み合わせた論文を1つ選択してください。

2. その論文の分野の中での位置づけを示しながら、論文の貢献を記述してください。
3. その論文の考えられる欠点を示し、もし可能なら論文の質を向上させる方策について述べてください。

## 11. 深層学習に基づく原子系シミュレータと材料開発への応用

1. 2017年以降に発表された論文で、機械学習を用いて原子系のシミュレーションを行うことを目的とした論文を2本選んでください。それら2本を比較しつつ論文の内容を説明してください。具体的には、それら論文で新しく提案された点や、互いの利点あるいは欠点などに注目しつつ客観的にまとめてください。
2. 機械学習を用いた原子系シミュレーションは発展途上にありますが、その中であなたが今後5年以内に解決されると考える課題をあげてください。課題1の論文で解決されていない欠点に注目しても構いませんし、そこから離れて材料探索に対する自分の意見を展開しても構いません。課題について現時点で達成されていない原因と、解決の糸口について自分なりの言葉でまとめてください。  
全体でA4, 2枚以内で記述してください(様式自由)。

## 12. ディープラーニングを使ったアニメーションへの応用

ディープラーニングを使ったアニメーションへの応用に関連する論文を一本選び、以下の点についてまとめてください。

1. 論文の要約
2. 技術的な貢献、新規性や応用可能性などに基づいて、この論文にご興味を持たれた理由
3. 仮にインターンシップでインパクトのある論文の執筆、もしくは実用的なアプリケーションの開発を行っていただくとします。この論文をどう活用したいか、また、そのために乗り越えなければいけない課題とその解決策  
全体でA4サイズ2枚以内で記述してください(様式自由)。

以下の国際会議で発表された論文から選ぶことを強く推奨します。

- CVPR 2019 <http://openaccess.thecvf.com/CVPR2019.py>
- ICCV 2019 <http://openaccess.thecvf.com/ICCV2019.py>
- CVPR 2020 (リストはまだ公開されていませんが、ネットで論文を公開している著者もいます。)

上記のリスト以外から論文を選ぶことも可能です。その場合、選んだ理由と、その論文を使ってアニメーションに関する課題をどう解決したいのかについて述べてください。

## 13. 工業プロセスのシミュレータ実装技術および周辺技術の開発

昨今 Jax [1] や DiffTaichi [2] といった微分可能プログラミングのフレームワークが提案されています。微分可能プログラミングを用いた物理シミュレーションとその設計に関する以下の問いに答えてください。

問1: 比較的新しい提案である DiffTaichi [2] の論文を解説してください。微分可能プログラミングの特徴を明らかにしながら、この論文の主張をあなたの言葉で要約してください。また、この提案の特色・欠点を明らかにしながら、物理シミュレーションに用いる観点で批評してください。

問2: 物理シミュレーションを実現するために微分可能プログラミングフレームワークが持つべき機能および設計について、あなたの考えを述べてください。以下に論点の例を示しますが、これにとらわれる必要はありません。

- 記述言語の設計：自動微分のための言語機構、並列化のための言語機構、物理的に意味があることを保証するための型システム、Perturbation Confusion の防止など
- コンパイラの設計：必要となるパス・中間表現など
- ランタイムの設計：自動微分のためのデータ構造の管理、並列化の実現方法、ガベージコレクションなど

関連研究を適切に引用してください。必要ならば図表を用いても構いません。図表や参考文献を含め、全体でA4用紙2ページ以内でレポートを作成してください。想定分量は問1・問2で各1ページずつです。

[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. 強化学習のロボティクスへの応用

1. ここ2年以内に公開された最近の論文で、強化学習のロボティクスへの応用に関する論文の一つを選んでください。
2. その論文の分野の中での位置づけを示しながら、論文の貢献を記述してください。
3. その論文の考えられる欠点を示し、もし可能なら論文の質を向上させる方策について述べてください。  
全体でA4, 2枚以内で記述してください(様式自由)。

## 15. 自動運転のためのコンピュータビジョン

1. ここ2年以内に公開された論文で、このテーマに関する内容で、あなたがインターン期間中に行いたい研究と関連する論文の一つを選んでください。つまり、自動運転に使えるような、segmentationやobject detection, trackingなどに関する論文を選んでください。
2. その論文の分野の中での位置づけを示しながら、論文の貢献を記述してください。
3. その論文の考えられる欠点を示し、もし可能なら論文の質を向上させる方策について述べてください。  
全体でA4, 2枚以内で記述してください(様式自由)。

## 16. Optuna 開発

以下の問題のうち、いずれか一つを選択し、解答してください。

[問題 A]

ハイパーパラメータ最適化では、最適化対象のハイパーパラメータのうち、ごく一部だけがモデルの性能に支配的な影響を与えている、というケースがしばしばあります。各ハイパーパラメータの重要度を評価するための既存手法の一つを記述してください。また、その手法の理論面ないし実用面での制限と、それへの対処方法を説明してください。

[問題 B]

Optuna (バージョン1.3.0) のRDBStorageが抱えている実行速度上のボトルネックや問題点を思いつく限り記述してください。

また可能であれば、その改善案も記述してください。必要であれば、RDBStorageのインタフェースを拡張したり、DBのスキーマを変更しても構いません。

## 17. 分散強化学習の技術開発

IMPALA [1] と SEED RL [2] を比較しながら、以下のことについて説明してください。

1. よくある画像分類の分散学習とこのような分散強化学習との違い
2. これらのシステムアーキテクチャとアルゴリズムについて

また、これらの問題点や改良点などについて自分なりの考察してください。

[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. Localizationに関する技術開発

1. Localizationに関連する技術の各テーマのなかで検討したいテーマを選択して、それに関連する論文を1つ選択して下さい（ここ2年以内に公開された論文である方が望ましい）。直接法/間接法ハイブリッドSLAM開発を選択された場合には以下の2つの論文のうちどちらかを選択ください。

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/lsdslam>

2. その論文の分野の中での位置づけを示しながら、論文の貢献を記述してください。
  3. その論文の考えられる欠点を示し、もし可能なら論文の質を向上させる方策について述べてください。
- 全体でA4, 2枚以内で記述してください(様式自由)。

## 20. 機械学習を用いた数量ファイナンス

### 課題

次の課題1と課題2のうちから1つを選んで解答してください。

なお課題1と課題2の間に優劣はなく、どちらかを選んだからといって、評価が良くなったり悪くなったりすることはありません。

レポートにおいては、10pt以上のフォントを用いてPDF形式・A4用紙1ページから2ページ程度で報告をしてください。また、そのレポートとは別に、作成において必要となったプログラムなどはレポートの後にPDF形式で付録として添付することも可能です。

### 課題1

以下のKenneth R. French教授のウェブサイトでは、ファクターと呼ばれる収益率のデータが配布されています。これらのファクターの中から好きなもの1つ以上を選んで、収益を上げるときに有用となる定性的な特徴・定量的な特徴を分析し報告してください。

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

### 課題2

以下の論文のうちから1つを選んで解説してください。

なお論文は第一著者の姓のアルファベット順に並んでいます。

特定の論文を選んだからといって、評価が良くなったり悪くなったりすることはありません。

また説明においては理論的側面よりも実応用の観点を重点的に説明してください。

[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-loser investment strategy using interpretable deep reinforcement attention networks. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019.
- [17] 貴方が好きな2020年5月1日時点で出版・公開されている論文 /Your favorite paper published as of May 1, 2020

## 21. 機械学習ベースの物理シミュレーションモデル開発

現実の物理現象を計算機上で再現・予測することについて、次の要素を含めながら、あなたの考えを述べてください。

1. あなたが計算機上で再現・予測したいと考える物理現象について、簡潔に説明してください。
2. その物理現象を科学理論に基づく数値計算によってシミュレーションする方法について説明してください。
3. その物理現象を予測するために機械学習技術を活用している既存の方法を1つ取り上げ説明してください。
4. その物理現象を計算機上で表現、再現、あるいは予測するために機械学習技術を用いることの意義、課題、および展望についてあなたの考えを述べてください。

関連研究を適切に引用してください。必要ならば図表を用いても構いません。図表や参考文献を含め、A4用紙2ページ以内でレポートを作成してください。想定分量は上記要素1-3について1ページ、要素4について1ページです。

## 22. 医用画像を用いた機械学習の種々の問題に関する研究

医用画像の解析には、施設・機器間差が大きいため均質なデータを集めづらいという課題や、アノテーション作業に必要な時間的・人的コストが高いため十分な数のアノテーション済みデータを集めづらいという課題があります。これらの課題（どちらか一方もしくは両方）を解決する為の技術に関連する論文をひとつまたは複数選び、論文における提案技術が課題に対してどのように貢献し得るか、その提案技術の改善点についてA4用紙1～2枚で論じてください。医用画像はX線画像やCT/MRI画像、病理画像など、臨床で用いられる画像全般を指します。

## 23. ディープラーニングの社会実装に向けた機械学習の研究

あなたは、とある機械学習系の会議の査読を引き受けました。以下の論文の中から、あなたが面白いと思うものを任意に1つ選び、擬似的な査読コメントを書いてください。

### Step1. 論文を選ぶ

以下の2つの国際会議で出版済みのもののなかから、論文をひとつ選んでください。

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

### Step2. 模擬査読コメントを書く

選んだ論文に対する査読コメントを書いてください。査読コメントでは、下記の設問に答えてください。

- (1) 選んだ論文のタイトル、著者、URL (例: <http://proceedings.mlr.press/v97/abcdefg19.html>)
- (2) この論文の貢献内容およびそのインパクトについて、サマリーを記述してください。
- (3) この論文の強みを3つ挙げてください。
- (4) この論文の弱みを3つ挙げてください。
- (5) 詳細なコメントを自由に記述してください。

- これまでの設問をふまえて、論文に対する中立的な批評を行ってください。
- accept/rejectの形式での結論や、著者に対する質問を書いたりする必要はありません。

A4用紙で1ページまたは2ページのPDFファイルを提出してください。フォントサイズは10pt以上としてください。

## 24. 機械学習を用いた数量ファイナンス

## 課題

次の課題1と課題2のうちから1つを選んで解答してください。

なお課題1と課題2の間に優劣はなく、どちらかを選んだからといって、評価が良くなったり悪くなったりすることはありません。

レポートにおいては、10pt以上のフォントを用いてPDF形式・A4用紙1ページから2ページ程度で報告をしてください。また、そのレポートとは別に、作成において必要となったプログラムなどはレポートの後にPDF形式で付録として添付することも可能です。

### 課題1

以下のKenneth R. French教授のウェブサイトでは、ファクターと呼ばれる収益率のデータが配布されています。これらのファクターの中から好きなもの1つ以上を選んで、収益を上げるときに有用となる定性的な特徴・定量的な特徴を分析し報告してください。

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

### 課題2

以下の論文のうちから1つを選んで解説してください。

なお論文は第一著者の姓のアルファベット順に並んでいます。

特定の論文を選んだからといって、評価が良くなったり悪くなったりすることはありません。

また説明においては理論的側面よりも実応用の観点を重点的に説明してください。

- [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] 貴方が好きな2020年5月1日時点で出版・公開されている論文 / Your favorite paper published as of May 1, 2020