

機械学習エンジニアコース

Week4 Session

－ 機械学習の基礎 －



DIVE INTO CODE

2020年7月30日(木)
富永 修司



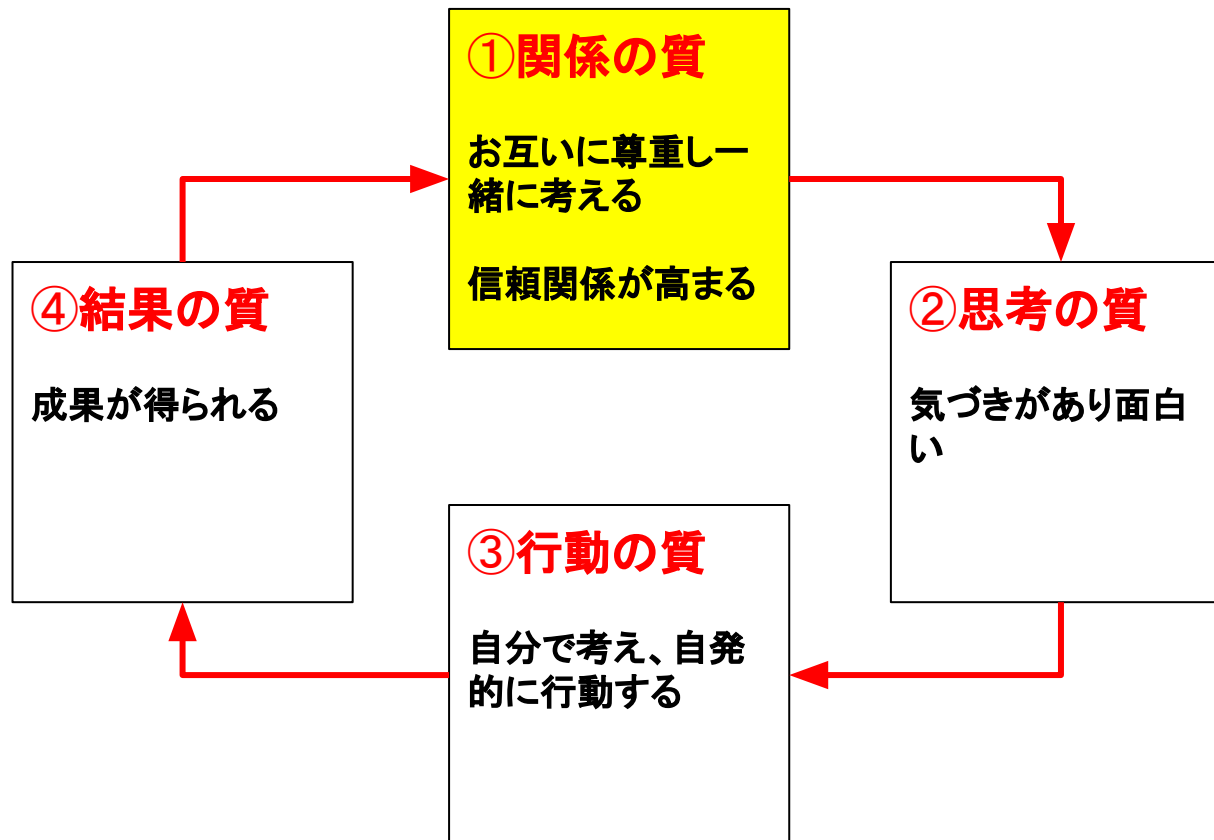
今日の流れ

1. チェックイン・KPT
2. 講義
3. お昼休み
4. ペアプログラミング
5. KPT・チェックアウト



チェックイン

同期生全員が、共に学びあう関係性ができており、具体的な思考や行動、結果につながる好循環ができている状態。

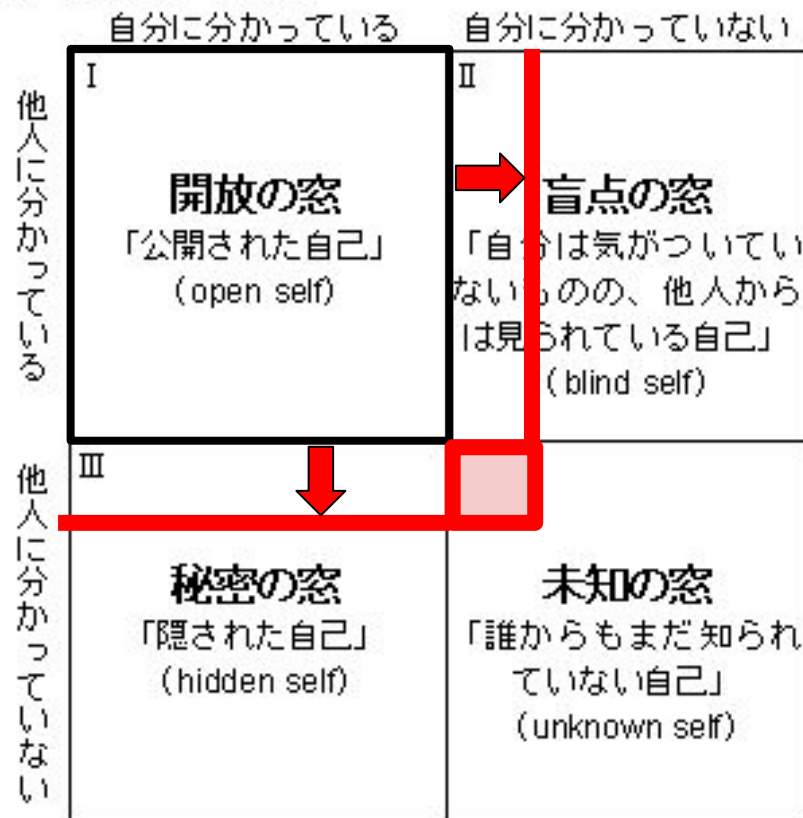


参考: 組織の成功循環モデル <https://jinjibu.jp/keyword/detl/815/>



チェックイン

ジョハリの窓の Open Self を広げる過程で気づきが生まれる。人は、必要性に”気づき”、それが学ぶキッカケとなる。



出典: フリー百科事典『ウィキペディア (Wikipedia)』<https://ja.wikipedia.org/wiki/%E3%82%B8%E3%83%A7%E3%83%8F%E3%83%AA%E3%81%AE%E7%AA%93>



チェックイン(一人1分程度で)

チェックインは**自分の状態を共有**する事です。

ペアプロをするにあたり自分の状態を共有する事、**周りの状況を
知る事**がより効果的な成果を生みます。

ペアプロの基本は**相互の関係の質**が重要になります。関係の質を
高める入り口は**相互のコミュニケーション**です。

- ・名乗り
- ・今日の気持ち・意気込み(喜怒哀楽、期待、緊張等)
- ・何か一言 etc →次の方を指名

↓
KPT



構成

1. 提言
2. 導入
3. 今日の目的
4. 授業前課題の確認
5. 授業課題



完璧を目指すよりもむしろ終わらせろ

Done is better than perfect.

マーク・ザッカーバーグ





カリキュラム

ゴールから逆算して設計されたカリキュラムになっています。数歩先を見据え、走りながら考えてください。

就職

機械学習エンジニアになる。

Term3(10月)

問題を定義して、時間内に解決できる。

Term2(9月)

現在の問題を認識し、既存の解決策を適用できる。

Term1(8月)

古典的理論を知り、定石を身につける。

事前学習(7月)

道具を活かす思考を身につける。



カリキュラム

Term3(10月)

問題を定義して、時間内に解決できる。

- 調査
- 仮説を立てる
- 条件を知る
- SQL
- データセット作成
- Docker
- Raspberry Pi
- 公開



カリキュラム

Term2(9月)

現在の問題を認識し、既存の解決策を適用できる。

- 深層学習
- 画像認識
- 自然言語処理
- 論文読解
- コードリーディング
- OSS
- フレームワーク



カリキュラム

Term1 (8月)

古典的理論を知り、定石を身につける。

- 教師あり学習
- 教師なし学習
- スクラッチ
- Kaggle



カリキュラム

事前学習(7月)

道具を活かす思考を身につける。

- プログラミング (Python)
- 機械学習のための数学
- 探索的データ分析
- 機械学習の基礎
- オブジェクト指向



導入 - 大切な考え方

今月は、道具を活かす思考を身につける。

	◎ Good	× Not Good
1	「何があればできるだろう」と自分に問う	「まだ習ってないからなあ」と立ち止まる
2	「本当にあっているのか」と疑う	「○○○に書いてあったから」と信じ込む
3	「まずはやってみよう」と手を動かす	「もっと分かってからやろう」と慎重になる



今日の目的

学びの目的。目的が、人の役割と必要な学びを明確にする。明確な学びは、成長実感と自信につながる。

	目的とすること	目的としないこと
1	仲間とプログラムの考え方を学ぶ	関数をたくさん覚える
2	機械学習の基礎を知る	課題を早く完成させる
3	新人ビジネスマンの気持ちになる	



今日の目的: **機械学習の基礎を知る**

「機械学習の基礎を知る」

そもそも機械学習の基礎とは何か。
モデル作りにこだわることではない。

- **機械学習の一連の流れの実装と実行を繰り返す**



今日の目的: 新人ビジネスマンの気持ちになる

「新人ビジネスマンの気持ちになる。」

分析ツールを使う上で大切にしたい姿勢。以下のようなイメージを持つてみる。

- 自分は新人ビジネスマン
 - ドメイン知識がない
 - データはある
- 会社で先輩や上司にホウレンソウする
 - 目的としているビジネスゴールがある



授業前課題の確認

授業前課題の解説を行います。

解説後は解説を踏まえ、授業前課題を元にペアプログラミングを実施し、お互いのコードをレビューして気付きを得ましょう。



授業課題

DIVER 授業前課題で身につけた力を活用して、より実践的な問題にチャレンジ！

1. 信用情報の学習

- a. コンペティション内容の確認
- b. 学習と検証
- c. テストデータに対する推定
- d. 特徴量エンジニアリング

参考情報

- HomeCredit_columns_description.csv
<https://www.kaggle.com/c/home-credit-default-risk/data>



授業課題

Kaggleコンペティションに取り組むフロー

1. 問題提起を理解する
2. 評価指標を理解する
3. PublicとPrivateの比率を確認する
4. EDA(データ分析)する
5. First Submissionする
6. 指標値の関数をつくる
7. 前処理する
8. Feature Engineering(特徴量エンジニアリング)する
9. Trainingする
10. 指標値で評価する
11. Submissionする



授業課題

【注意事項】HomeCredit_columns_description.csv を見て、データセットの列についての説明を確認しよう。

Data (688 MB)

Data Sources

application_test.csv 48.7k x 121
application_train.csv 308k x 122
bureau.csv 1.72m x 17
bureau_balance.csv 27.3m x 3
HomeCredit_columns_description.csv
HomeCredit_columns_... 219 x ...
installments_payme... 13.6m x 8
POS_CASH_balance... 10.0m x 8
previous_applicatio... 1.67m x 37
sample_submission.... 48.7k x 2

	A	B	C	D	E
1	≡	Table	Row	Description	翻訳
2	1	application_{train_test}_id	SK_ID_CURR	ID of loan in our sample	私たちのサンプル中のローンのID
3	2	application_{train_test}_id	TARGET	Target variable (1 - client with payment difficulties)	ターゲット変数 (1 - 支払が困難なクライアント: 彼/彼女は)
4	5	application_{train_test}_id	NAME_CONTRACT_TYPE	Identification if loan is cash or revolving	識別ローンは、現金または回転であれば
5	6	application_{train_test}_id	CODE_GENDER	Gender of the client	クライアントの性別
6	7	application_{train_test}_id	FLAG_OWN_CAR	Flag if the client owns a car	フラグクライアントが車を所有している場合
7	8	application_{train_test}_id	FLAG_OWN_REALTY	Flag if client owns a house or flat	旗クライアントは、家やフラットを所有している場合
8	9	application_{train_test}_id	CNT_CHILDREN	Number of children the client has	子供の人数クライアントが持っています
9	10	application_{train_test}_id	AMT_INCOME_TOTAL	Income of the client	クライアントの収入
10	11	application_{train_test}_id	AMT_CREDIT	Credit amount of the loan	ローンの与信額
11	12	application_{train_test}_id	AMT_ANNUITY	Loan annuity	ローン年金
12	13	application_{train_test}_id	AMT_GOODS_PRICE	For consumer loans it is the price of the goods for sale	消費者ローンのために、それは、ローンが与えられるため、
13	14	application_{train_test}_id	NAME_TYPE_SUITE	Who was accompanying client when he was applying for loan	彼は融資を申請されたとき、誰が、クライアントに付随して
14	15	application_{train_test}_id	NAME_INCOME_TYPE	Clients income type (businessman, working, married, etc.)	クライアントの収入の種類 (実業家、作業、産休、...)
15	16	application_{train_test}_id	NAME_EDUCATION_TYPE	Level of highest education the client achieved	最高の教育の水準クライアントを実現
16	17	application_{train_test}_id	NAME_FAMILY_STATUS	Family status of the client	クライアントの家族の状況
17	18	application_{train_test}_id	NAME_HOUSING_TYPE	What is the housing situation of the client (renting, own, etc.)	クライアントの住宅事情は何である (親と同居、レンタル、
18	19	application_{train_test}_id	REGION_POPULATION_RELATIVE	Normalized population of region where client lives	クライアントが住んでいる地域の正規化された人口 (高い数
19	20	application_{train_test}_id	DAYS_BIRTH	Client's age in days at the time of application	塗布時の日中のクライアントの年齢
20	21	application_{train_test}_id	DAYS_EMPLOYED	How many days before the application the person was employed	何日前に申請人は、現在の雇用を開始しました
21	22	application_{train_test}_id	DAYS_REGISTRATION	How many days before the application did client register	何日の適用前に、クライアントは自分の登録を変更しました
22	23	application_{train_test}_id	DAYS_ID_PUBLISH	How many days before the application did client publish	何日の適用前に、クライアントは、彼がローンを適用したと
23	24	application_{train_test}_id	OWN_CAR_AGE	Age of client's car	クライアントの車の時代
24	25	application_{train_test}_id	FLAG_MOBIL	Did client provide mobile phone (1=YES, 0=NO)	クライアントは、携帯電話を提供しました (1 = YES, 0 = NO)

Googleスプレッドシートで **GOOGLETRANSLATE関数** を使おう！
<https://support.google.com/docs/answer/3093331?hl=ja&authuser=0>



(再掲)今日の目的

学びの目的。目的が、人の役割と必要な学びを明確にする。明確な学びは、成長実感と自信につながる。

	目的とすること	目的としないこと
1	仲間とプログラムの考え方を学ぶ	関数をたくさん覚える
2	機械学習の基礎を知る	課題を早く完成させる
3	新人ビジネスマンの気持ちになる	



完璧を目指すよりもまず終わらせろ

Done is better than perfect.

マーク・ザッカーバーグ





チェックアウト(一人1分程度で)

一日の学習お疲れ様でした！

最後に気持ちを共有して解散しましょう！

- ・名乗り
- ・今の気持ち・振り返り(喜怒哀楽、期待、緊張等)
- ・何か一言 etc →次の方を指名



宿題

「Sprint 機械学習スクラッチ入門」

次回の授業までに提出しましょう！^^
授業後にDIVERにて紐付けをします。

機械学習の基礎 完