会計データで機械学習を実践しよう（応用編）

講師紹介

- 野呂祐介
- 経歴
  - 2016年3月　慶應義塾大学理工学部　卒業
    脳信号（fMRI）データ解析（多次元時系列の分析）
    2016年3月　慶應義塾大学理工学研究科　入学
  - 2017年11月　公認会計士試験合格
  - 2018年3月　慶應義塾大学理工学研究科　修了
  - 2018年3月　有限責任監査法人トーマツ　入所
    金融事業部にて、リース会社やクレジットカード会社の監査、監査のためのデータ分析
  - 2021年2月　株式会社ローソンデジタルイノベーション　入社
    施策効果測定、クーポンや広告のターゲティング、需要予測
- その他
  - 自然言語処理の研究や財務諸表データやKAMデータの分析を趣味でやっています。

本講義の目標
『機械学習を業務に活かすイメージを掴む』

「機械学習でどんなことができるのか」を知っていれば、

- ➢ プロトタイプは生成AIに作ってもらい、試すことができる。
- ➢ 事例を見た時に、裏側がイメージできるようになる。

※アルゴリズムの解説は省略して、道具としての使い方を中心にお伝えします。（どの手法もそれを実現させるアルゴリズムに支えられています。）

本講義の目標
『機械学習を業務に活かすイメージを掴む』

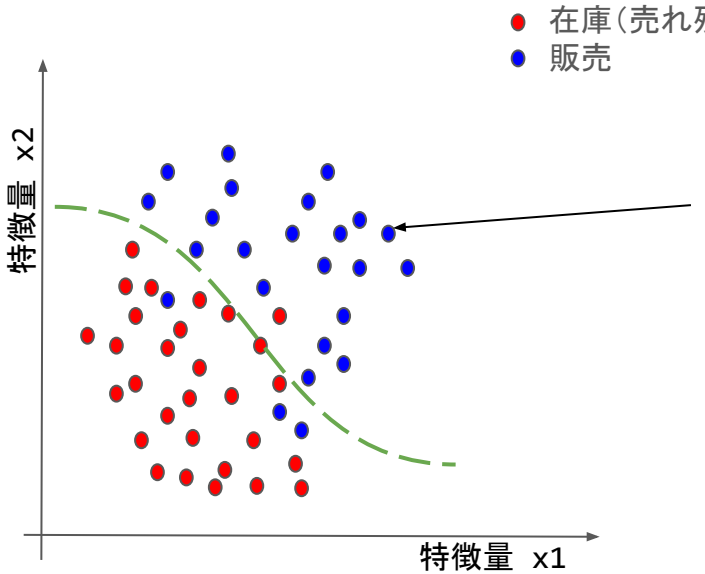
- 4.1 機械学習の実践（説明10:40～11:00、演習・質問 11:00～11:10）
  - 教師あり学習の復習とPythonによる実践（生成AIを用いたプログラミング学習）
- 4.2 機械学習アウトプットの使い方と解釈（説明11:10～11:45、演習・質問 11:45～12:00）
  - 評価指標、説明可能AI、点予測、区間予測、分布予測、予測スコア
  - モンテカルロシミュレーション、反実仮想の予測、効果測定
  - 説明可能AI、教師なし学習
  - Pythonによる機械学習の実践
- 4.3 データに応じた柔軟な機械学習（12:00～12:10）
  - 異常検知、半教師あり学習、自己教師あり学習、転移学習、大規模言語モデル、 In context learning、傾向スコア、ドメイン適応、強化学習

各章で扱うキーワードを記載しています。

4.1 機械学習の実践

機械学習とは？

機械学習モデル：データの異質性に注目し、ばらつきのパターンを活かして、複雑な仮説を表現。



データひとつひとつを何にするか（異質性を見出す単位）は様々です。
→分析結果の解釈したい単位やアクションをとりたい単位で決めます。

- 時間（会計期間）
- 商品・サービス
- 店舗・営業所
- 取引先・顧客
- 契約・取引

機械学習の適用余地

機械学習モデル：データの異質性に注目し、ばらつきのパターンを活かして、複雑な仮説を表現。

⇨関心のある対象が、たくさんの構成単位に分けられると機械学習の適用余地があります。

関心のある対象	構成単位	機械学習の適用余地
減損損失	全国にある200店舗	◯
売上	巨大船舶売却取引が2つ	???
売上	10万人の消費者向けサブスクリプションサービス	
ホテル（不動産）の評価	100部屋が稼働中	
ある事業の資産グループの評価	3年の中期経営計画を評価	

機械学習の適用余地

機械学習モデル：データの異質性に注目し、ばらつきのパターンを活かして、複雑な仮説を表現。

⇨関心のある対象が、たくさんの構成単位に分けられると機械学習の適用余地があります。

関心のある対象	構成単位	機械学習の適用余地
減損損失	全国にある200店舗	◯
売上	巨大船舶売却取引が2つ	△
売上	10万人の消費者向けサブスクリプションサービス	???
ホテル（不動産）の評価	100部屋が稼働中	
ある事業の資産グループの評価	3年の中期経営計画を評価	

機械学習の適用余地

機械学習モデル：データの異質性に注目し、ばらつきのパターンを活かして、複雑な仮説を表現。

⇨関心のある対象が、たくさんの構成単位に分けられると機械学習の適用余地があります。

関心のある対象	構成単位	機械学習の適用余地
減損損失	全国にある200店舗	◯
売上	巨大船舶売却取引が2つ	△
売上	10万人の消費者向けサブスクリプションサービス	◯
ホテル（不動産）の評価	100部屋が稼働中	???
ある事業の資産グループの評価	3年の中期経営計画を評価	

機械学習の適用余地

機械学習モデル：データの異質性に注目し、ばらつきのパターンを活かして、複雑な仮説を表現。

⇨関心のある対象が、たくさんの構成単位に分けられると機械学習の適用余地があります。

関心のある対象	構成単位	機械学習の適用余地
減損損失	全国にある200店舗	◯
売上	巨大船舶売却取引が2つ	△
売上	10万人の消費者向けサブスクリプションサービス	◯
ホテル（不動産）の評価	100部屋が稼働中	◯
ある事業の資産グループの評価	3年の中期経営計画を評価	???

機械学習の適用余地

機械学習モデル：データの異質性に注目し、ばらつきのパターンを活かして、複雑な仮説を表現。

⇨関心のある対象が、たくさんの構成単位に分けられると機械学習の適用余地があります。

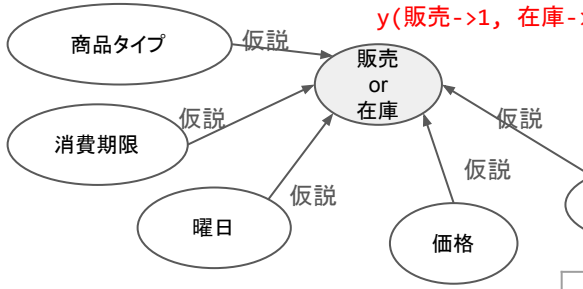
関心のある対象	構成単位	機械学習の適用余地
減損損失	全国にある200店舗	◯
売上	巨大船舶売却取引が2つ	△
売上	10万人の消費者向けサブスクリプションサービス	◯
ホテル（不動産）の評価	100部屋が稼働中	◯
ある事業の資産グループの評価	3年の中期経営計画を評価	◯

機械学習は何をしているのか

機械学習モデルは複雑な仮説を表現します。例えば、ある商品が当日売れるかを予測したいとします。

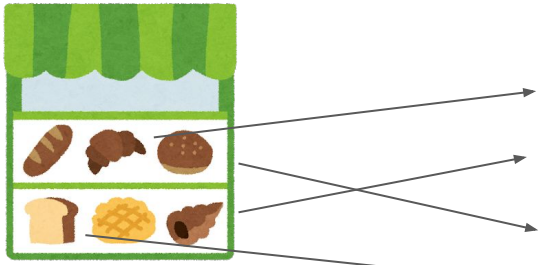
説明変数：x

目的変数：
y(販売->1，在庫->0)



青線の部分が大きいほどyは1に近づく

$$y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \text{商品タイプ} + \beta_2 \text{消費期限} + \beta_3 \text{曜日} + \beta_4 \text{価格} + \beta_5 \text{在庫数})}}$$



目的変数　説明変数

商品番号	販売or在庫	商品	商品タイプ	消費期限	曜日	価格	在庫数
A_1	販売	A	チョコ系	2024/5/2	火	150	50
B_1	在庫	B	チョコ系	2024/5/3	火	200	50
C_1	販売	C	ピザ系	2024/5/2	火	120	50
C_2	在庫	C	ピザ系	2024/5/1	火	120	50

月刊 企業会計 連載「Pythonではじめるまでの会計データマネジメント」-分析指向の戦略的データ収集 -図表1を改編

機械学習は何をしているのか

どのβを大きくするかによって、様々な仮説を表現

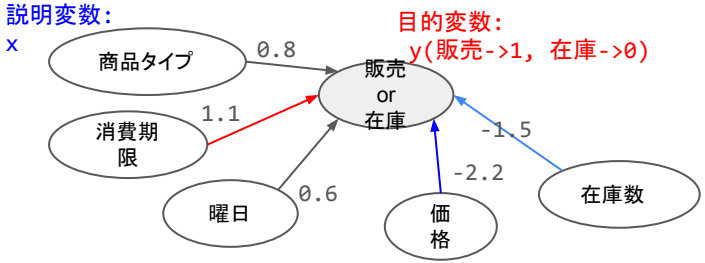
$$y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \text{商品タイプ} + \beta_2 \text{消費期限} + \beta_3 \text{曜日} + \beta_4 \text{価格} + \beta_5 \text{在庫数})}}$$

訓練データ（2024/4/1～2024/5/1）

目的変数　　説明変数

商品番号	販売or在庫	商品	商品タイプ	消費期限	曜日	価格	在庫数
A_1	販売	A	チョコ系	2024/5/2	火	150	50
B_1	在庫	B	チョコ系	2024/5/3	火	200	50
C_1	販売	C	ピザ系	2024/5/2	火	120	50
C_2	在庫	C	ピザ系	2024/5/1	火	120	50

モデルがデータに当てはまるように（目的変数を当てられるように）、パラメータを調整することで（訓練）、有力な仮説を提案します。



最も目的変数を当てられるβの組み合わせ（有力説）を推定

$$y = \frac{1}{1 + e^{-(0.2 + 0.8 \text{商品タイプ} + 1.1 \text{消費期限} + 0.6 \text{曜日} - 2.2 \text{価格} - 1.5 \text{在庫数})}}$$

機械学習は何をしているのか

$$y = \frac{1}{1 + e^{-(0.2 + 0.8\text{商品タイプ} + 1.1\text{消費期限} + 0.6\text{曜日} - 2.2\text{価格} - 1.5\text{在庫数})}}$$

予測対象データ（2024/5/2）

目的変数　　説明変数

商品番号	販売or在庫の予測	商品	商品タイプ	消費期限	曜日	価格	在庫数
A_101	販売	A	チョコ系	2024/5/4	木	150	50
B_101	在庫	B	チョコ系	2024/5/3	木	200	50
C_101	販売	C	ピザ系	2024/5/2	木	120	50
C_102	販売	C	ピザ系	2024/5/3	木	120	50

有力な仮説を反映したモデルにより、目的変数を予測します。

機械学習は何をしているのか

予測対象データ（2024/5/2）

目的変数

番号	販売or在庫の予測
A_101	販売
B_102	在庫
C_103	販売
C_104	販売

- 予測結果をアクションに応じてアクション: 朝の在庫よりも予測販売数が大幅に多い場合には追加発注、少ない場合には値引き。（会計データサイエンスではモニタリングに応用されることが多いです）
- シミュレーション: 例えば5/1に割引キャンペーンを行った時に、実績ーシミュレーション値で効果を測る。（会計データサイエンスでは見積もりに応用されることが多いです）
- 仮説から学ぶ:（分析用のデータセットがあるが、予測対象データがない場合には、ナレッジを活かす他ありません）

実践

生成AIを使用したプログラミング学習

生成AIを使用したプログラミング学習の強みは、みなさんのコードややりたいことをベースに解説を作ってくれることです。

生成AIによるプログラミング学習のポイント

- プロンプト（生成AIへの入力）は、テンプレートを用意して、編集は最小限にします。（生成AIへの指示に時間がかかっては本末転倒）
- プロンプトは、用途が決まれば、オプション設定のようなものです。（テンプレートは編集しやすく、変更管理がしやすいようにします。）
- コード生成は用途が決まっている場合（インプットとアウトプットが決まっている関数）に適しています。用途が変更する可能性がある場合、後々変更に対応できるような仕様をプロンプトで指示することが難しいです。また、途中から要件を追加していくと沼にはまる場合が多いです。
- 現状のでは、エラーの修正は簡単なものでないと沼にはまってしまうことがあります。（※AIエージェント等で今後改善の見込みあり）
- 説明ドキュメントを書かせる精度が高いため、コードを教材に学習を進めやすいです。

※Google colaboratolyでは、”AI で生成”機能を有効にすると有効にしたGoogleアカウントではノートブックのデータをモデルの学習に追加されてしまうため、秘密情報や個人情報は書かないでください。

生成AIを使用したプログラミング学習（例）

あなたは入門者向けpythonプログラミング学習の補助アシスタントです。

サンプルデータについて
python変数dataのカラムは以下の通りです。
  'sold_today':bool,# 目的変数
  'date':str, #仕入日付 YYYY/MM/DD
  'prod_id_unique':str, # 商品番号 (棚に並んでいる商品ひとつひとつを区別 ex A_1_20240401, A_2_20240401, ...)
  'product_name':str, # 商品名 {A,B,C,D<E}
  'expiry_date':str, # 消費期限 YYYY/MM/DD
  'product_type':str, # 商品タイプ {チョコ,ピザ,食パン,クロワッサン}
  'price':str, # 価格 100～250
  'weekday':str, # 曜日 {月, 火, 水, 木, 金, 土, 日}
  'weather':str, # 天候 {晴れ, 雨, 曇り}
  'same_prod_type_stock':str # 同じ商品の開始在庫数

指示
変数dataのsold_todayカラムを予測するlogistic回帰モデルのpythonコードを提供してください

次に注意してください。
(- サンプルデータを作成してください。)
- コードブロックを実行しながら挙動を確認できるように、できるだけ関数化しないでください。
- 入門者がわかるような易しいコメントをつけてください。
- 各変数がどのようなものなのかコメントしてください
- これを教材として、pythonの各変数のデータ形式やクラスやメソッドを解説してください
- コメントは日本語で書いてください

はい、承知しました。以下にpythonのコードを生成します。

プロンプトに扱うデータの情報を入れる（カラム名、型）。

型は迷ったらstr（文字型）を指定しておく。（後で変換できるため）

生成AIを使用したプログラミング学習（例）

あなたは入門者向けpythonプログラミング学習の補助アシスタントです。

datanのsold_todayカラムを予測するlogistic回帰モデルのpythonコードを提供してください

次に注意してください。
- コードブロックを実行しながら挙動を確認できるように、できるだけ関数化しないでください。
- 入門者がわかるような易しいコメントをつけてください。
- 各変数がどのようなものなのかコメントしてください
- これを教材として、pythonの各変数のデータ形式やクラスやメソッドを解説してください
- コメントは日本語で書いてください

以下に続けてコードを記載してください。

```
import pandas as pd
filename = "/content/sample_data_pads/dataset/store_dataset.csv"
dtypes = {
   'sold_today':bool,# 目的変数
    'date':str, #仕入日付
    'prod_id_unique':str, # 商品番号 (棚に並んでいる商品ひとつひとつを区別 ex A_1_20240401, A_2_20240401, ...)
    'product_name':str, # 商品名 {A,B,C,D<E}
    'expiry_date':str, # 消費期限 YYYY/MM/DD
    'product_type':str, # 商品タイプ {チョコ,ピザ,食パン,クロワッサン}
    'price':str, # 価格 100～250
    'weekday':str, # 曜日 {月, 火, 水, 木, 金, 土, 日}
    'weather':str, # 天候 {晴れ, 雨, 曇り}
    'same_prod_type_stock':str # 同じ商品の開始在庫数
}
data = pd.read_csv(filename, index_col=None, dtype=dtypes, encoding='utf-8')
```

はい、承知しました。以下にpythonのコードを生成します。

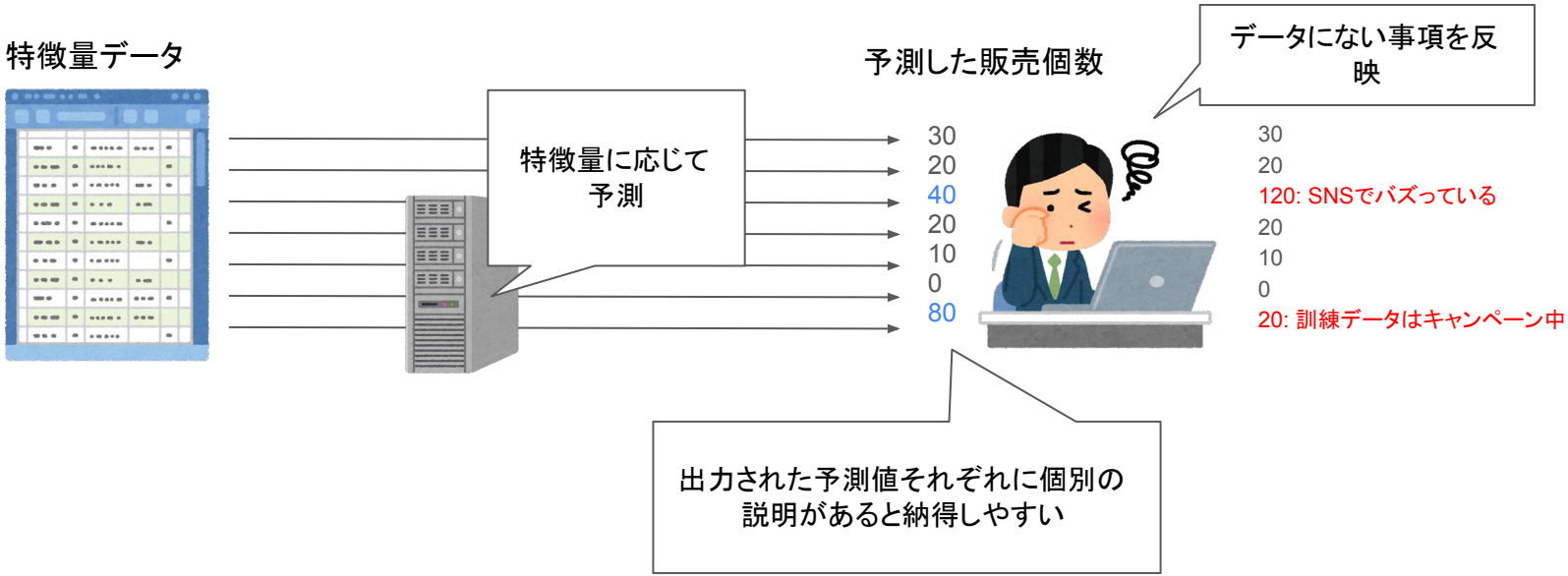
データの読み込み部分は書いて与えて、続きを生成してもらう。

4.2 機械学習アウトプットの使い方と解釈

機械学習の予測結果に応じてアクション

会計業務はステークホルダーへの説明や監査対応のように人間に対して行われることが多く、また、責任も伴います。こうした場合、機械学習で業務を完全に自動化するよりも、機械学習と人間が共同して業務を行う方が導入がしやすいです。

機械学習と共同で意思決定をする人間は、①**機械学習モデルがどのような性質を持ち** 、②**どのようなデータで訓練し** 、③**予測精度がどの程度であるのか** 把握している必要があります。



①機械学習モデルの性質

機械学習モデルによっては、機械学習モデルが表現できる仮説の範囲を限定したり、特定の仮説を選好するバイアスがあります（帰納バイアス）。

また、パフォーマンスを発揮しやすい訓練データの量にも違いがあります。

事前知識や取得できるデータに合わせた機械学習モデルを選ぶことが重要です。

モデル	性質（モデルの数式や推定アルゴリズム）
線形回帰モデル	特徴量の大きさに比例して増加減少する。
ロジスティック回帰	0～1の間で、特徴量が大きくなるほど1に近づくが、そのスピードは0に近づいていく。
LightGBM	少ない特徴量で精度を上げることを重視。
ワイブル回帰モデル	時間（変数）を大きくすると0に収束していく。
深層学習モデル	バイアスが少ない。（大量のデータがないとパフォーマンスが発揮しづらい）

②機械学習モデルの訓練データのバイアス

機械学習モデルの訓練データには少なからずバイアスが入ります。

例

- データの取得期間
  - 特定の時期に集中しやすく、キャンペーンや経済状況、季節の影響がある。
- サンプルの選択バイアス
  - 契約顧客のみのデータを使っているため、潜在顧客に当てはまるかわからない。
  - 上場企業のみのデータを使っているため、非上場企業、特別目的会社、個別店舗に当てはまるかわからない。
- データの観測のバイアス
  - 在庫がなければ販売データがない（需要が測れない）。
  - 会計認識されなければ観測されない。
  - 減損認識がなければ事業価値が再評価されない。

③機械学習モデルの信頼性（精度評価）

機械学習モデルの生成する予測値、予測区間（予測分布）、異常検知における正常範囲、説明可能AIによる分析結果、効果測定等の全ては、予測精度のもとで（信頼性の限界のもとで）解釈可能となります。

評価指標の解釈はタスクや業務要求水準で変わりますが、一つの尺度として

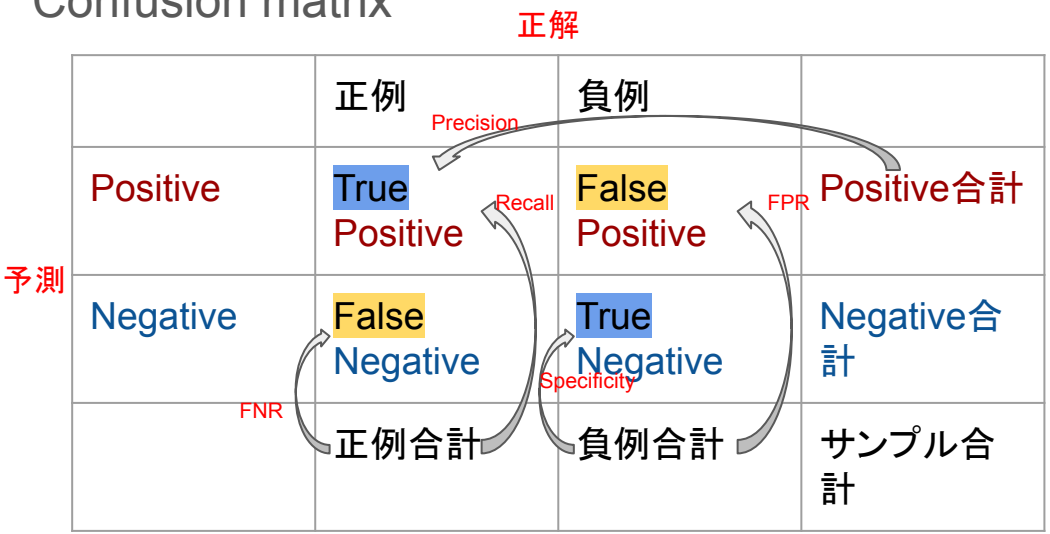
- ランダムに予測した場合よりも大幅な精度向上があるか
- 簡単な予測をした場合よりも精度が向上するか

また、間違ったサンプルが別の手段で対策可能か（エラー分析）の観点も重要です。

タスク	評価指標の例
点予測	（予測と実績の）平均絶対誤差、平均二乗誤差
2値分類	False Positive Rate（偽陽性率）、AUC（予測スコアによる分離能力）
ターゲティング	Precision（優良顧客の的中率（打率））、Recall（優良顧客の取りこぼしのなさ）
異常検知	Recall（過去の異常の取りこぼしの少なさ）、過去の異常例を何番目に検出できるか

【補足】2値分類モデルの評価指標

Confusion matrix

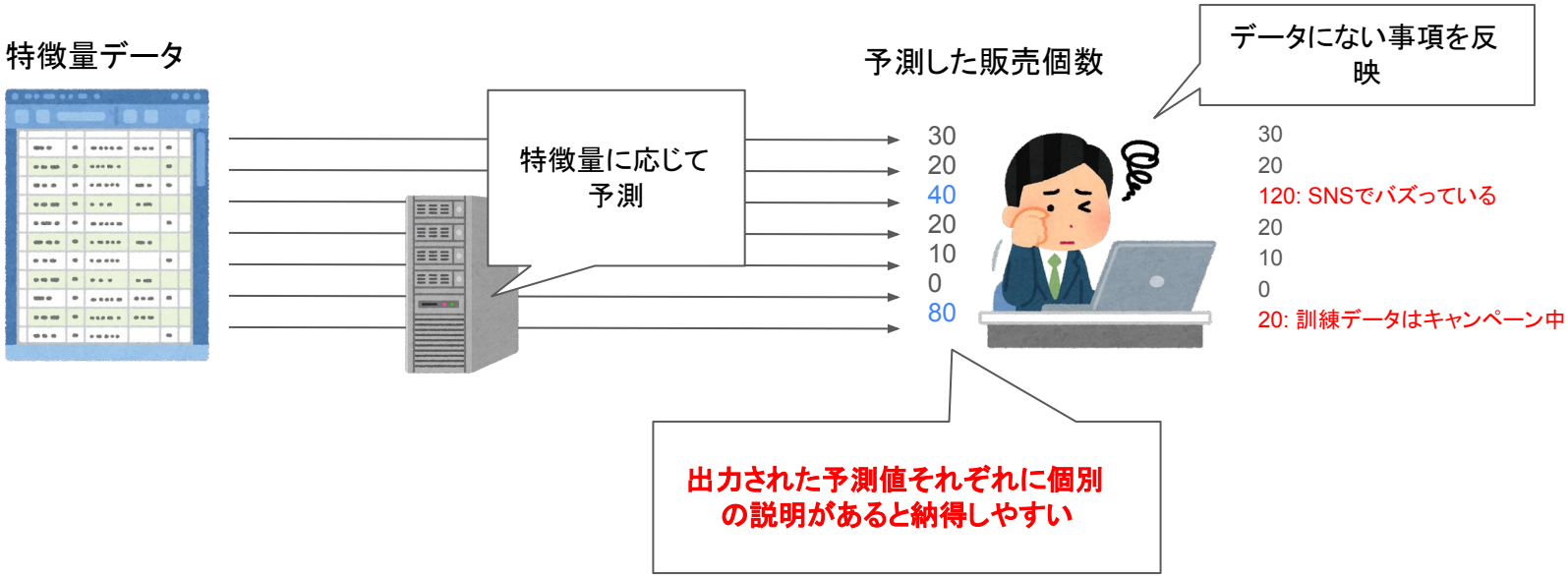


正解率(Accuracy)	(TP+TN)/サンプル合計	TP、FP、FN、TNのうち特に重視するがない場合に使います
適合率(Precision)	TP/Positive合計	打率と考えるとわかりやすい、効率性の指標
再現率(Recall) 感度(Sensitivity) TruePositiveRate	TP/正例合計	取りこぼしていないか、仮説検定における検出力(1-β)
偽陽性率 FalsePositiveRate	FP/負例合計	仮説検定におけるType 1 Error(α)
特異度(Specificity) TrueNegativeRate	TN/負例合計	仮説検定における(1-α)
偽陰性率 FalseNegativeRate	FN/正例合計	仮説検定におけるType 2 Error(β)

機械学習の予測結果に応じてアクション

会計業務はステークホルダーへの説明や監査対応のように人間に対して行われることが多く、また、責任も伴います。こうした場合、機械学習で業務を完全に自動化するよりも、機械学習と人間が共同して業務を行う方が導入がしやすいです。

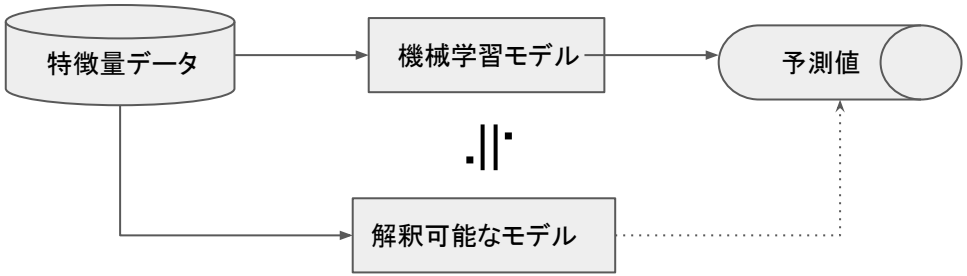
機械学習と共同で意思決定をする人間は、①**機械学習モデルがどのような帰納バイアスを持ち** 、②**どのようなデータで訓練し** 、③**予測精度がどの程度であるのか** 把握している必要があります。



説明可能AI

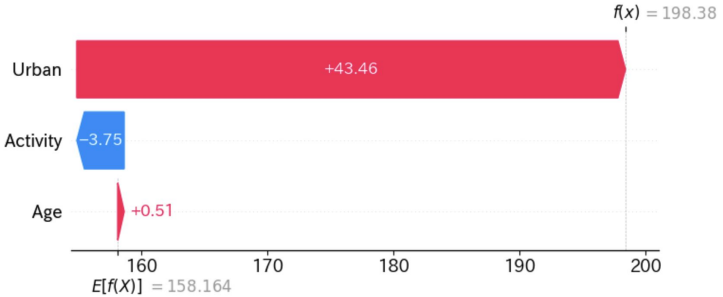
機械学習モデルは特徴量と目的変数の関係を柔軟に学習し、高い精度で予測できるが、解釈が困難です。一方で、オペレーションに人間と機械学習モデルの両方が入る場合には、機械学習の出力を人間が解釈（納得）するための予測の根拠が欲しいことが多いです。

そこで、複雑な機械学習を解釈可能なモデルで近似し、機械学習モデルを説明するモデルを作成します。



- 与信結果の説明
- 発注量の説明（店舗の発注担当者の判断材料）
- 機械学習モデルの監査

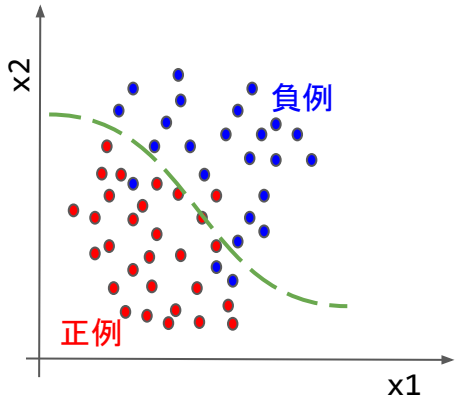
SHAPという手法では機械学習モデルの予測値を説明変数ごとの貢献度の和で表せます。

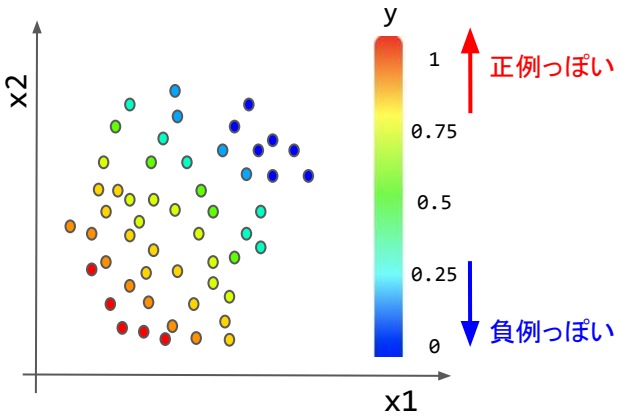


分類を予測スコアで行う

正例と負例の違いを学習し、特徴量 xについてそのサンプルが正例か負例かを予測します。(正例を 1、負例を0に対応させることで、目的変数が1か0を取る回帰とみなせます)

特徴量xに対して正例らしさのスコアを予測することもできます。(目的変数が 0から1の値を取る回帰)



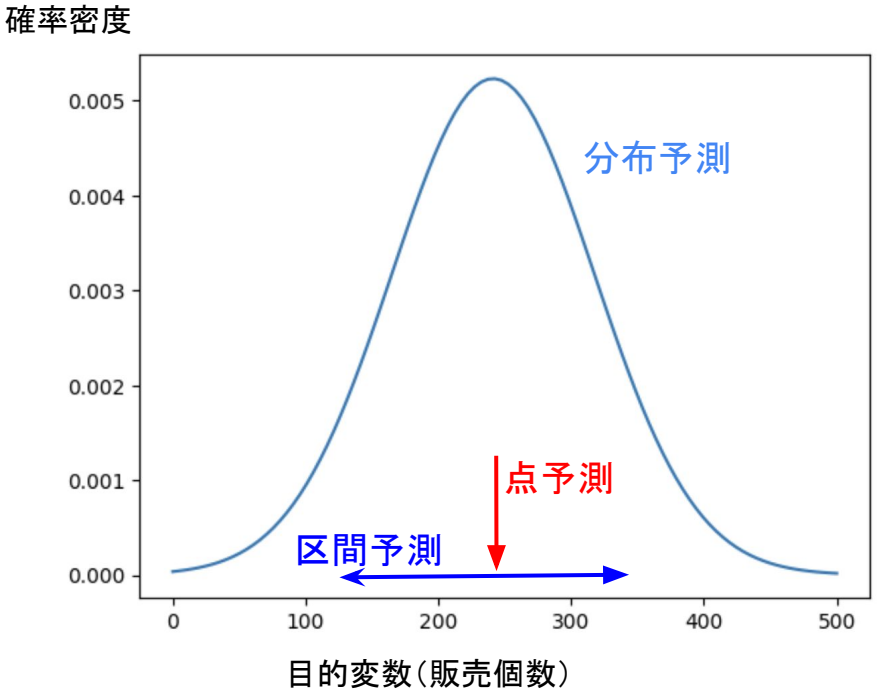


- 販売スコアがワースト50の商品に値引きシールを貼る
- 貸倒予測スコア(信用スコア)が高い取引先を特別に調査・管理対象とする
- 広告クリック予測スコアが上位の顧客へのPUSH通知広告
- 会計監査におけるリスクアプローチ

【補足】点予測・区間予測・分布予測

機械学習による回帰では、目的変数を予測値だけでなく、その範囲や分布としても予測できます。

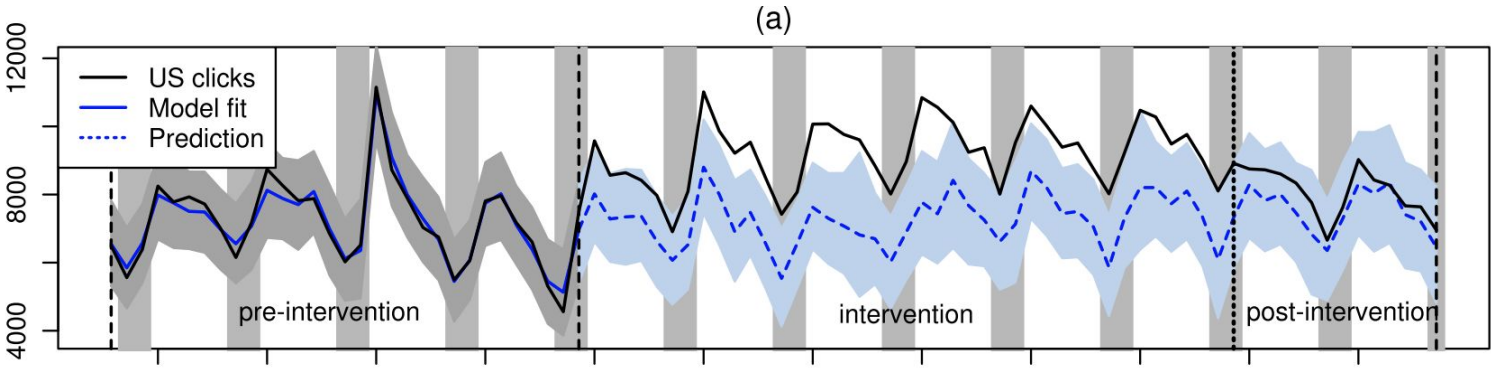
- 発注のための販売個数の区間（分布）予測
- 会計監査における分析的実証手続の監査人の許容範囲として80%予測区間を採用する



効果測定への機械学習応用（反実仮想の予測）

介入による効果の**平均**= ①介入を行った場合の**平均**ー②介入を行わなかった場合の**平均**

介入前の観測値、そのトレンドや周期性等から②を予測します。

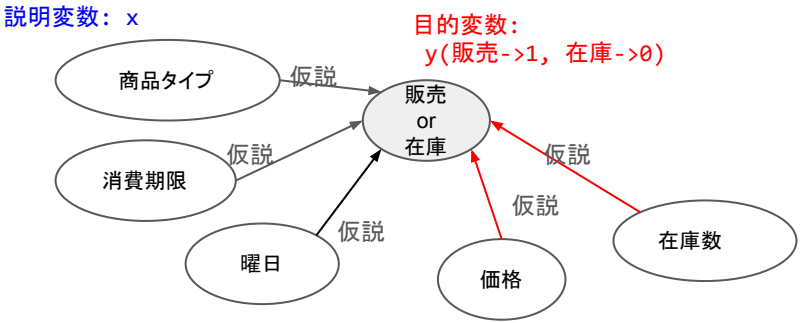


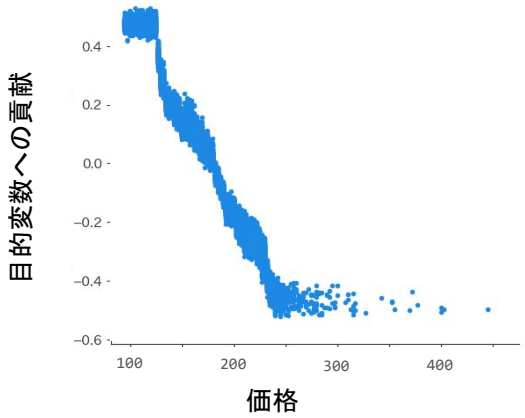
Kay H. Brodersen. Fabian Gallusser. Jim Koehler. Nicolas Remy. Steven L. Scott. "Inferring causal impact using Bayesian structural time-series models." Ann. Appl. Stat. 9 (1) 247 - 274, March 2015. https://doi.org/10.1214/14-AOAS788

説明可能AIによる知識発見

説明可能AIは機械学習モデルの予測結果に対して判断根拠を提示（推定）します。

さらに、十分に精度の高い機械学習モデルであれば、特徴量の組み合わせと目的変数の関係や特徴量と目的変数の非線形な関係も学習できている可能性があります。そのモデルを説明可能AIで解析することで特徴量と目的変数の関係の仮説発見ができる場合があります。（この仮説は独立したデータで検証することが望ましいです。）





4.2 Pythonによる機械学習の実践

4.3 データに応じた柔軟な機械学習
（少数データ、偏りのあるデータ、行動の学習）

ラベル付きデータがない場合（教師なし学習）

訓練データ（2024/4/1～2024/5/1）

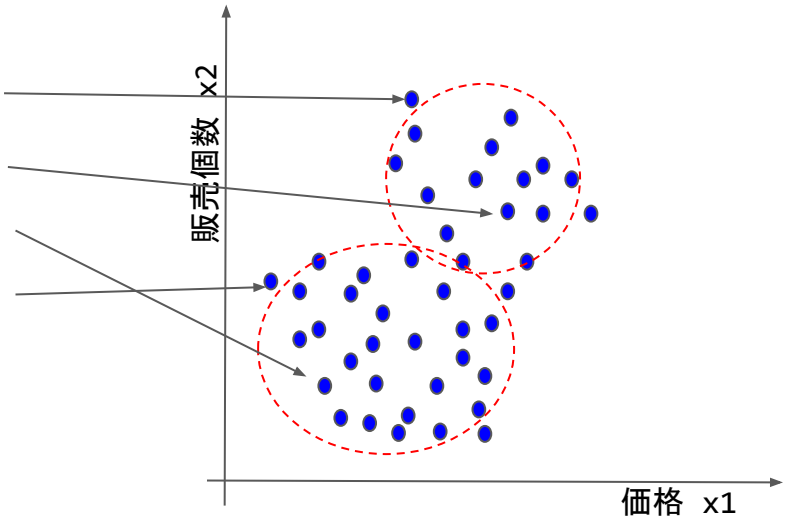
説明変数

商品	商品タイプ	消費期限	曜日	価格	販売個数
A	チョコ系	2024/5/2	火	150	200
B	チョコ系	2024/5/3	火	200	140
C	ピザ系	2024/5/2	火	120	120
D	ピザ系	2024/5/1	火	120	90

教師なし学習

特徴量同士の関係を学習。教師データがなくても使えます。

➔　グループ分け、特徴量の支配的なパターンの探索

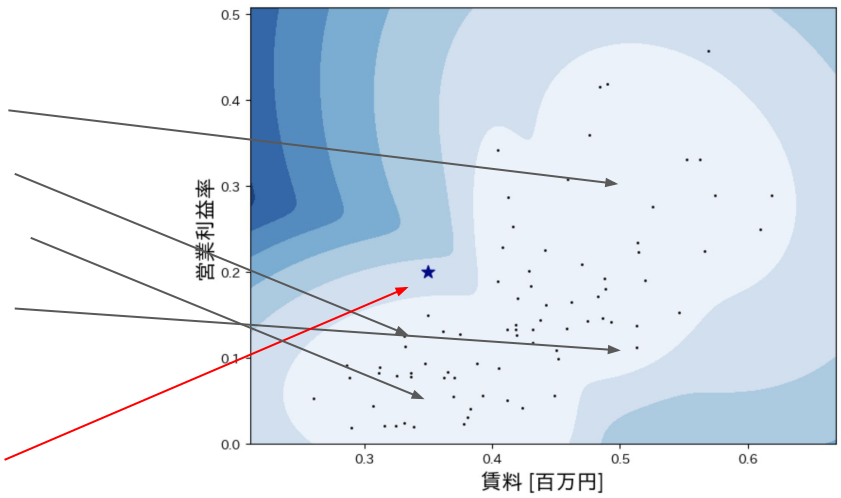


正例と負例のどちらかしかない場合（異常検知）

正常なデータの分布する範囲を機械学習で予測し（教師なし学習）、その分布から外れているデータを異常とするアプローチです。異常サンプルがある場合には正常と異常の分類タスクとして教師あり学習でモデルをつくることができます。

店舗番号	正常or異常	営業利益率	賃料
1	正常	0.3	520,000
2	正常	0.12	510,000
3	正常	0.11	340,000
4	正常	0.05	350,000

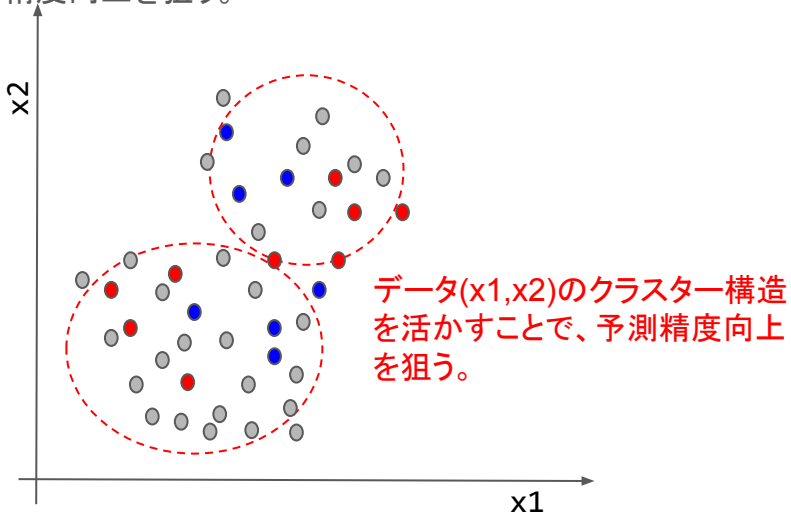
店舗番号	正常or異常	営業利益率	賃料
1	??	0.18	360,000



ラベル付きデータが少ない場合（半教師あり学習・転移学習）

半教師あり学習

教師データが少ない時に、豊富なデータで教師なし学習を行い精度向上を狙う。



類似のアプローチ

転移学習：異なるタスク・ドメインの教師があるデータで学習し、そのモデルを本命の教師ありデータで学習します。

- ➢ 一般的な文章で学習させたモデルを法律分野の文章で追加学習する。

自己教師あり学習：データから教師信号を定義して、教師あり学習を行います。（転移学習と組み合わせて使います）

- ➢ 自然言語処理で次に続く文章を予測する。
- ➢ 画像処理で回転させたり並行移動させただけの画像を同じ画像と判別させる。

【補足】大規模言語モデルを用いたIn-Context-Learning

大規模言語モデルはWeb上の膨大なデータを学習した結果、言語表現だけでなく、記載内容も学習しています。学習した記載内容の豊富な知識を活かす使い方があります。

商品番号	販売or在庫	商品	商品タイプ	消費期限	曜日	価格	在庫数
A_1	販売	A	チョコ	2024/5/2	火	150	50
B_1	在庫	B	チョコ	2024/5/3	火	200	50
C_1	販売	C	ピザ	2024/5/2	火	120	50

商品番号	販売or在庫	商品	商品タイプ	消費期限	曜日	価格	在庫数
A_101	??	A	ピザ	2024/5/4	木	180	50

アウトプットである営業終了時点での状態 {販売,在庫}を予測します。
次のサンプルでわからない部分はあなたの常識的な知識ベースで補ってください。

サンプル
インプット: 商品=A, 曜日= 火, 商品タイプ=チョコ, 価格=150
アウトプット: 販売

インプット:商品=B, 曜日=火, 商品タイプ=チョコ, 価格=200
アウトプット: 在庫

インプット:商品=C, 曜日=火, 商品タイプ=ピザ, 価格=120
アウトプット: 販売

問題
インプット:商品E, 曜日: 水, 商品タイプ: ピザ, 価格: 180
では商品Eの営業終了時点の状態はなんでしょうか。
回答フォーマット
アウトプット: (販売 or 在庫)

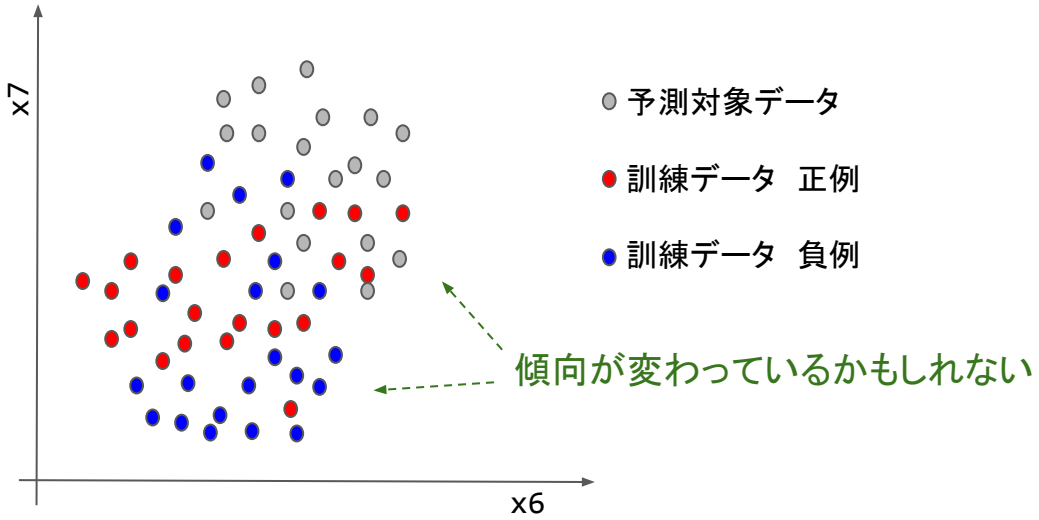
では、ステップバイステップで考えていきましょう。

はい。それでは、...

Vacareanu et al., 2024. From Words to Numbers: Your Large Language Model Is Secretly A Capable Regressor When Given In-Context Examples.https://arxiv.org/abs/2404.07544

訓練データと予測対象データの性質が異なる場合（ドメイン適応）

いくつかの特徴量において、予測値との関係が学習データと予測対象のデータで変わっているケースがあります。
予測対象に近いデータで学習させたいものの、予測対象に近いデータに絞るとサンプルサイズが小さくなってしまいます。→予測対象に近いデータを重視して学習するアプローチがとれます。

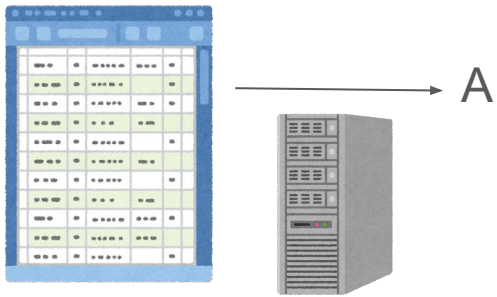


正解データがないが行動の結果が得られる場合（強化学習）

特徴量データ

正解はBです

特徴量データ



まあまあな予測です（5点/10点）

教師あり学習

予測した人件費に対して、正しい人件費が与えられます。その差が小さくなるように、モデルを調整できます。

強化学習（即時フィードバック）

スロットマシンが2台あり、合計n回でたくさん当てたいとします。予測して選んだ台が回すべき台であるかはわかりません。一方で、その回当たったかどうかがわかります。

人間にとってわかりやすい、倫理的に良い文章を出力するように調整するタスクを考えます。どのような文章を生成すべきか（正解）はわかりません。でも、生成してみた文章の（人間による）採点結果がわかります。

強化学習（遅延フィードバック）

将棋のように数ある打ち手の中から予測して選んだ手に対して、どの手が正解だったかはわかりません。選んだ手が良い手であるかも直ぐにはわかりません。勝敗が決まったときに初めてわかります。

【まとめ】本講義の目標
『機械学習を業務に活かすイメージを掴む』

会計データサイエンスとは、

統計や機械学習（**数学**）とプログラミング（**IT**）スキルを使って、**ビジネス**の課題を解決するプロセスであり、特に会計課題や会計データを扱うもの。

本講義では**数学・IT**の実践例をいくつか紹介しました。みなさんの身の回りの**ビジネス**で「自分の周りでこういう問題なら解けそう」「こういうデータがあればできそう」など応用を考えてみてください。

ご清聴ありがとうございました

お気軽にご連絡ください！

X(twitter): @nororororororor

PyCPA: https://join.slack.com/t/pycpa/shared_invite/zt-4sqhf39x-vkA9hVAt4aJDLXkOU7waNQ