『Python3/Rによるデータサイエンス概論: 線形回帰分析の理論・開発』

株式会社セラク
SI本部
スマートソリューション部
データサイエンティスト
柴田 怜

キーワード:
線形単回帰分析、ガウス分布、最小二乗法、決定係数、直接相関と疑似相関

要旨

1. この特論計画・進行形態を俯瞰する。
2. 統計理論を概説し、開発演習(Python3/R)を出題する。
3. 最も単純な例題として線形単回帰分析を用いる。
4. 様々な演習を以て技術習熟に努められたい。
5. データサイエンスの開発未経験者に向け、Kaggleを紹介する。
6. 開発時の参考資料及び要領を提示する。
7. 次回の特論に用いる指定教本を提示する。
8. 更に学びたい方に向け、確率論・統計数学(概念・数式・専門)の推奨書籍を紹介する。

目次

研究概要

項目	概要
学位論文	『エネルギー改革策が及ぼした環境・経済・社会的影響: 日独英仏国の実証分析と国際比較』
テーマ	気候変動対策が及ぼす環境・経済・社会影響の国際比較・実証分析
要旨	当時、我が国は、先の原子力発電所事故並びに気候変動の深刻化に鑑み、特にドイツの再生可能エネルギー普及策に倣った。 『エネルギー基本計画』(日本国)は、3E+S(Energy, Environmental Conservation, Economy growth and Safety)を基本方針にエネルギー改革策を施した。 この政策評価として、ドイツ・フランス・イギリスの先行事例に対し、環境・経済・社会影響を基準に和英文の先行文献を以て国際比較・実証分析を行った。 その結果、我が国のエネルギー改革策は必ずしも成功していなかったと結論を下した。 以上にを踏まえ、環境税を財源とする更なる気候変動対策を第二次安倍内閣の新・経済政策における投資先に提言した。
定量分析手法	対数線形重回帰分析、主成分分析、分散分析、微積分法、パス図描画、プロット分析
但書	1. この研究は、2016年度までの事象に鑑みた結果である。 2. ESG(Environmental, Social, and Corporate Governance)投資等の最新動向を踏まえていない。

定量分析手法の開発経験

方法	開発経験(R/Python3/VBA)
計量時系列分析	ARIMA(Autoregressive Integrated Moving Average)モデル、VAR(Vector Auto Regressive)モデル、状態空間モデル、確率ボラティリティ変動モデル、偏グレンジャー因果性検定、ベイズ動的線形モデル、パネルVARモデル、多変量確率ボラティリティ変動モデル、Dynamic-SEM(Structural Equation Modeling)、構造VARモデル、動学的応用一般均衡モデル、DICE(Dynamic Integrated Climate-Economy)モデル、差分の差分法
機械学習	k-means、判別分析、SVM(Support Vector Machine)、主成分分析、ランダムフォレスト、ニューラルネットワーク、ディープラーニング、深層強化学習、LightGBM、XGboost、t-SNE、ノンパラメトリックベイズ、密度準拠型クラスタリング、画像解析
数理最適化	ゲーム理論、動的計画法、整数線形最適化、グラフ最適化、(非)線形最適化、制約あり最適化
数値解析	幾何ブラウン運動、常微分方程式、偏微分方程式、乱数シミュレーション
統計的有意差検定	無相関検定、カイ二乗検定、*t*検定、分散分析、U検定、線形回帰分析、コンジョイント分析
統計モデルとベイズ推定	一般化線形(混合)モデル、MCMC(Markov chain Monte Carlo methods)、因子分析、構造方程式モデル、交差検証法、ブートストラップ法、AIC(Akaike information criterion)、評価関数
実験計画法	分散分析表、一元配置分散分析、二元配置分散分析、線形モデルと最小二乗法
空間統計学	空間的自己相関、空間重み行列、クリギング、時空間統計解析
産業連関分析	LCA(Life Cycle Assessment)、波及効果、雇用係数、線形計画法、応用一般均衡モデル

凡例

1. 開発経験は前提とせず、適時、開発演習を交えて理論を解説する。
2. 個別の定量分析手法については、次回以降の特論で扱うものとする。
3. 定量分析手法の理論を理解し、開発演習を行うことで、内容理解を伴う開発技術(データサイエンス)の習熟を到達目標とする。
4. この特論資料及びソースコードは、講義後に共有する。
5. 開発言語は、原則としてPython3、必要時にR又はExcel VBAを用いる。
6. IDE(Integrated Development Environment)は、Jupyter Notebookとする。
7. Python3、R及びJupyter Notebookを事前にインストールされたい。
8. 総合演習の解答は、s.shibata@edixweb.jpに提出されたい。
9. この提出ファイルは、査読した後、講評を併記して返却する。

第1部 イントロダクション

(疑問)

1. データサイエンスの実例は、どのようなものか?
2. データサイエンティストに将来性はあろうか?
3. 一方、開発言語は、それぞれどのように異なろうか?

1. データサイエンスの社会実装

分野	内容	事例
Webマーケティング	ネット販売履歴から、顧客がどのような商品に関心がありそうかを分析し紹介する。	ネット通販サイトである商品を購入したり検索したりすると現れる「この商品を購入した人は、こんな商品にも関心をもっています。」
アクセスログ解析	インターネットのログ履歴のデータを分析し、イベントチケットの価格を変更して販売する。	YouTubeで「坂本冬美」を聞くと脇に「石川さゆり」がお勧めとして出る。
画像解析	医療機器画像の解析技術を用いて得られたデータをベテラン医師レベルで診断装置が高速かつ高精度で判断する。	CT画像、MRI画像及び内視鏡データから癌を検出する。

(参考)
日本最大級のAI専門メディアのAINOWは、データサイエンスの様々な社会実装に関する実例や業界動向を日々アップデートしている。

2. データサイエンティストの将来性

1. そもそも出来る方が限られ、就職・転職しやすい。
2. 人月単価が高い。
3. 年々、案件の増加傾向が見込まれる。

- AI(人工知能)
- RPA(Robotic Process Automation)
- 一般社団法人 データサイエンティスト協会の設立

3. 開発言語の比較表

(引用:『前処理大全』P.11)

	SQL(DataBase)	R	Python(3系)
動作環境	データベース上 (メモリ上にのらないデータサイズも扱える)	Rプロセス上 (基本的にはメモリ上にのるデータサイズを扱う)	Pythonプロセス上 (基本的にはメモリ上にのるデータサイズを扱う)
記述量	多い	やや少ない	普通
処理速度	早い	やや遅い	早い
分散処理	適切なSQLを書けば自動で行われる	可能ではあるが、記述コストが大きくかかる	可能ではあるが、記述コストが大きくかかる
システム化環境	充実している	あまり充実していない	充実している
計算機能	一部関数で実現されているのみ、機械学習はほぼ不可	充実している	充実している
描画機能	なし	充実している	充実している

Excel VBA:

1. Excelの作業効率向上を主な目的とし、データ分析には必ずしも向いていない。
2. しかし、Excelを用いずに済むIT業務は殆どない為、習得する価値はある。

第2部 概論

(問題意識)

- この特論を以て何を得ようか?
- データサイエンスとは何か?
- データサイエンティストには、どのようなスキルセットを要しようか?
- 経験上、何を言えようか?

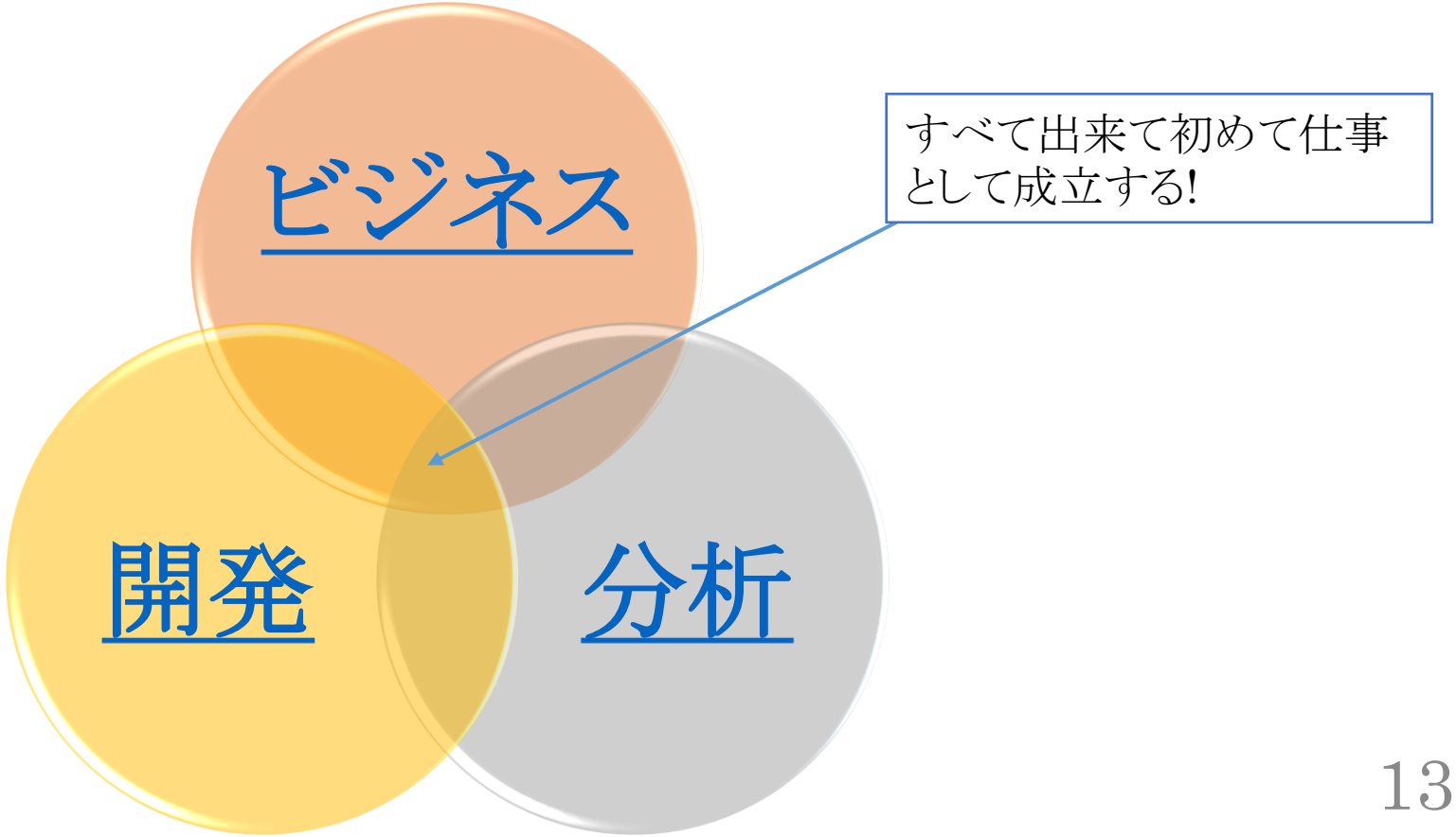
1. 目的・意義

技術習熟	具体例	想定案件(例)
定量分析手法の理論・開発	・機械学習 ・計量時系列分析 ・経営工学 ・実証分析 ・効果検証	- データサイエンティスト - 定量分析系の研究開発
開発言語(R/Python)	・基本文法 ・ライブラリ ・関数定義 ・前処理 ・グラフ描画 ・AI	- Python開発技術者 - 機械学習エンジニア - データサイエンティスト(開発技術も必須となる。)

2. データサイエンスとは

(定義)

あるビジネス課題に対し、既存の様々なデータに前処理及び探索的データ解析を行った後、定量分析手法の開発によって実証結果を得て顧客向けの説明を行う技術である。



3. ビジネス・スキルとは

代表例	目的・意義	構成(例)
リポーティング	(1)分析結果を顧客向けに説明する。 - 数式を利用は、可能な限り控える。 - 顧客が理解したいのは、問題に対する結果とその根拠である。 - プレゼンテーション資料にそのまま利用する為、PowerPointを用いることが多い。	(正副標題) 1. 目的 2. 問題の所在 3. 方法 (1)定量分析手法の要旨 (2)選定理由の説明 4. 結果 ・グラフ ・図表 5. 考察 Appendex A-1. 専門用語 A-2. 定量分析手法の詳細説明 A-3. 作動環境 A-4. 実行手順 A-5. ソースコードのリンク A-6. 引用統計 A-7. 参考文献
プレゼンテーション	(2)その不明箇所に質疑応答する。 - どのように丁寧に説明しても、顧客には必ず不明箇所が生じる。 - 想定質問は、予め別シートに模範解答を用意する方法もある。 - 本当に答えられない場合は、降参して後日調査して報告する。	

4. 開発・分析

フェーズ	目的	例
要件定義	- 何をいつまでにどのように開発・分析するかを明確化する。 - その達成に至るまでのロードマップ作成に寄与する。 - 顧客と齟齬を起こさない為、契約書等の定量的な再現性を担保することを推奨する。	- アウトプットの定義 - 期日の設定 - 開発言語 - 定量分析手法 - DB(Database)
前提情報		- 過去案件情報のキャッチアップ - スピードと正確さのどちらをより重視するか?
前処理	集積したローデータを開発・分析に即した形に加工する。	6~9月の集計期間に店舗の売上高を抽出する。
定量分析手法 **の検討・開発**	- 機械学習 - 計量時系列分析 - 統計的有意差検定	来月、この店舗の来店人数はどのように変化するかを予測する。
視覚化	- 分析結果を直感的に理解可能に伝える。 - リポーティングの根拠資料として利用する。	- ダイアグラム描画 - 回帰分析表
リポーティング	- 問題に対する解、その分析結果及び考察を論述する。 - 顧客の要望に応じ、導出(確率論・統計数学)を後述する。	- 分析結果報告書 - 仕様書

5. 経験談

No.	項目	説明
1	前処理	1. データサイエンスにおける作業時間の8割を占めると言われる。 2. 実際には、その前段階の要件定義に曖昧な箇所が散見されたり、前提情報を充分に説明されなかったりする。 3. その結果、無意味な手戻りが生じることに問題の所在があるケースが目立つ。 4. したがって、要件定義や前提情報に曖昧な点を残すと、後々苦労することになる。
2	設計	1. 何をどのように行うかのロジックを考えないまま、知っている関数や構文を組み合わせても解決しない。 2. 設計してから、開発・分析を行うことを推奨する。
3	統計理論	1. 統計理論を理解せずにライブラリを読み込み、数字を代入するのみでは、顧客向けの説明において、内容を説明することができず、必ず失敗する。 2. したがって、「統計理論 → 開発言語」の順序で技術習熟することを推奨する。

6. 総括

概論	説明		
結論	データサイエンスとは、経験的に得たデータを基に、様々な定量分析手法に係るシステムを設計・開発し、社会的・学術的に有益な知見を求める為の体系的技術		
用途	(例)	定量分析手法	業務
	予測	勾配ブースティング決定木	機械学習を用いた営業向け販促資料 半導体監視センサー値の閾値開発 小売店における来店人数予測
	分類	ロジスティック回帰分析	迷惑メールの自動検知
	因果推定	偏グレンジャー因果性検定 差分の差分法 構造方程式モデリング	インシデントの原因究明
	実証比較	統計的有意差検定	新薬の副作用に関する実証研究
	次元圧縮	t-SNE k-means 密度準拠型クラスタリング	KPIの抽出 次元の呪いの克服

第3部 ケーススタディ

(基本方針)

1. この特論の進行形態を説明する為、線形単回帰分析を例示する。
2. 統計理論の概説及びその数理補足を行う。
3. 設計演習において開発手順を考案されたい。
4. その後、Python3/Rの開発演習を行う。
5. 技術論文を課題とする。
6. 仕上げたら、s.shibata@edixweb.jpへ提出されたい。
7. 査読・講評して返却する。
8. 特に優秀な技術論文の著者には、プレゼンテーションを依頼する。

1. 線形単回帰分析とは

統計理論	概説
目的	ある目的変数に対し、ある説明変数がどの程度の直接相関係数を有するかを実証する。
定義	一般式:= $Y = \alpha + \beta X + \varepsilon$, Y: 目的変数の確率ベクトル ⊂ ガウス分布, α: 定数項, β: 直接相関係数 ⇔ 一変数導関数の微分係数, X: 説明変数の確率ベクトル, ε: 誤差項, 確率ベクトル: ある確率分布に従属する統計値の数列.
前提	この目的変数がガウス分布に従属しており、その説明変数に疑似相関がない。
方法	散らばっている各値の誤差を最小化する直線を引き、その回帰式を記述する。
手順	(1)目的変数と説明変数のローデータを収集・整理する。 (2)要約統計量を求め、ガウス分布に従属しているか否かを目視確認する。 (3)散布図を描画し、回帰直線を引くことで、相関関係に仮説を立てる。 (4)最小二乗法(OLS: Ordinary Least Squares)を以て回帰式を導出する。 (5)この回帰式の評価指標として決定係数を導出する。 (6)回帰分析表を出力し、その結果を要約する。

2. ガウス分布とは

確率論	概説	
定義	データが平均値の付近に集積するような独立な多数の因子の和として表される確率分布	
性質	平均値 = 中央値 = 最頻値	
根拠	中心極限定理	
補足	中心極限定理	大数の法則によると、ある母集団から無作為抽出した標本の平均は標本の大きさを大きくすると母平均に近づく性質
	大数の法則	試行回数を増加するにつれ、統計的確率は数学的確率に近似していく性質
	統計的確率	試行結果から求めた確率
	数学的確率	同様に確からしさを基に計算して求める確率
	同様に確からしさ	起こり得るすべての結果のいずれが起こる可能性もすべて同じとする性質
影響	*t*分布や*F*分布等の様々な確率分布の考え方の基礎になっているだけでなく、実際の統計的推測においても、仮説検定・区間推定等の様々な場面で利用される。	
実例	大学入試偏差値:= データの値を平均50、標準偏差10に正規化したときに示す指標	
別名	標準正規分布	

3. 最小二乗法とは

統計数学	概説
目的	線形回帰分析の解法となる。
定義	測定で得られた数値の組を適当なモデルから想定される一次関数・対数曲線等のある関数を用いて近似するとき、想定する関数が測定値に良い近似となるように残差平方和を最小とする係数を決定する方法
残差平方和	1. 残差とは、各データの平均値からのズレである。 2. 残差平方和は、この残差を二乗し、総和を取る。 3. 絶対値を取ってモデル指標とする為、残差を二乗する。
問題の所在	非線形性に対応することができない。
解決手段	1. 一般化線形化モデル 2. 非線形回帰分析
例	限界消費性向:= ケインズ経済学における家計による消費支出を表す関数
導出	ケインズ型消費関数:= $C = C0 + cY$, C: 消費, $C0$: 独立消費, c: 限界消費性向 ⇔ $dC/dY = c$, Y: 国民所得 ⇔ 実質GDP.

4. 決定係数とは

統計数学	概説
目的	線形回帰分析のモデル指標として利用する。
定義	目的変数の観測値に対する目的変数の予測値の説明力を表す指標
性質	0から+1.0までの値を取り、+1.0に近いほど分析が有効である。
問題の所在	重回帰分析においては、説明変数が増えるほど決定係数は+1.0に近似する。
解決手段	自由度修正済み決定係数を用いる。
重回帰分析	説明変数が複数ある線形回帰分析である。

5. 直接相関と疑似相関

論理学	直接相関	疑似相関
定義	因果関係の推定可能な相関関係	偶然に確認した相関関係
例	限界消費性向	見せかけの回帰

(但書)

1. 計量時系列分析や効果検証において、詳細に扱うことを計画している。
2. 現時点で理解不十分でも問題ない。

6. 導関数と微分係数

高等数学	導関数	微分係数
前提	ある関数の従属変数を独立変数について微分した際の概念	
定義	その結果、得た別の関数	接線の傾き
数式	$y = f(x) \Rightarrow y` = f(x)`$	$y = a + bx \Rightarrow dy/dx = b$

(但書)

1. 高等数学及びその開発(Python)は、第3回にて扱うことを計画している。
2. 更に学びたい方は、確率論・統計数学(概念レベル)を一読されたい。

7. 設計演習

(想定)

ある中学校に在籍する生徒52名の英語と数学の成績に鑑み、

数学の学習に伴う論理的思考力の向上が英文読解力に影響を及ぼすと、

その教職員(数学)は仮説を立て、線形単回帰分析を以て実証分析の開発を試みるものとする。

(設問)

この教職員(数学)が行うべき開発手順を設計・論述せよ。

8. 解答例(開発手順)

(1)目的変数と説明変数のローデータを収集・整理する。
・この場合、英語の成績と数学の成績を収集し、一つの.csvに整理する。
・数学が英語に及ぼす影響を実証する為、英語を目的変数、数学を説明変数とする。
(2)要約統計量を求め、ガウス分布に従属しているか否かを確認する。
①線形回帰分析は、目的変数がガウス分布に従属していることを前提とする。
②目的変数(英語の成績)について要約統計量を求める。
③「平均 = 中央値 = 最頻値」に近似しているかを目視確認する。
(3)散布図を描画し、回帰直線を引くことで、相関関係を目視確認する。
・必要なライブラリや関数を用い、開発効率を向上する(以下、同じ)。
・目視確認により、相関関係に仮説を立てる。
(4)最小二乗法(OLS: Ordinary Least Squares)を以て回帰式を導出する。
(5)この回帰式の評価指標として決定係数を導出する。
(6)回帰分析表を出力し、その結果を要約する。

9. 開発演習(Python3/R)

1. 先の例題を示すローデータを以下のリンクよりダウンロードせよ。
2. 前頁の開発手順を基に、Python3及びRにて再現せよ。

- 模範解答は、事前に用意してあるが、独自開発してから参考とせよ。

第4部 総合演習

(前文)

1. 今回の特論を仕上げる為、下記の総合演習を課題とする。
2. すべて同一の.zipに纏め、私宛に提出せよ。

(総合演習)

1. 確率論・統計数学
2. 技術論文
3. プレゼンテーション

1. 確率論・統計数学

(設問)

1. 最小二乗法を導出せよ。
2. 決定係数を導出せよ。
3. ガウス分布の性質を論証せよ。
4. 線形単回帰分析において、直接相関を用いる理由を論証せよ。

2. 技術論文

(設問)

「任意の命題に線形単回帰分析をPython3又はRにて開発し、ダイアグラムを描画しつつ、データサイエンスについて論述せよ。」(本文3000字程度)

(構成)

1. 目的・意義
2. 定義
3. 背景と問題の所在
4. 内容
5. 長所短所
6. 実例
7. 結果
8. 考察
9. Appendix(例: 実証結果、ソースコードのリンク等)

3. プレゼンテーション

1. 査読において特に優秀と看做した技術論文については、その著者に15分程度で結果を発表する場を次回の冒頭に設ける。
2. 事前配布資料は任意とする。
3. PowerPointの利用を可とする。
4. 事前配布資料及び発表資料は、s.shibata@edixweb.jpへ予め提出せよ。

Appendix

次頁以降は、今回の特論における説明に対し、下記を以て補論とする。

1. 専門用語の定義
2. 追加問題(Kaggle)とその解答例
3. 次回特論に向けた指定教本の紹介
4. 確率論・統計数学(概念・数式・専門)の推薦書籍
5. 参考(SQL/R/Python/VBA)
6. プログラミングの要領

A-1. 専門用語

専門用語	定義
勾配ブースティング決定木	学習器にあまり高性能なものを使わずに、弱分類器という感じで、予測値の誤差を新しく作った弱学習器がどんどん引き継いでいきながら誤差を小さくしていく方法
ロジスティック回帰分析	従属変数が二進法で記述され、ロジット分布に従うとき、重回帰分析に用いる一般化線形モデル
偏グレンジャー因果性検定	3つ以上の時系列データについて、あるデータ列を使うことで、ある別ののデータ列単体でそのデータ列の未来を推定するより良く推定する関係に基づく因果推定手法
差分の差分法	ある施策を行った際、施策を変数に含む群と含まない群で、それぞれ回帰分析を行い、その施策の因果推定を行う手法
構造方程式モデリング	潜在変数を仮定して行う重回帰分析であり、主に因果推定に用いられる。
統計的有意差検定	特定の代表値の比較において、所与の確率の下、差があるか否かを実証する手法
t-SNE	高次元データを2次元又は3次元に変換して可視化するための次元削減アルゴリズム
k-means	クラスタの平均を用い、与えられたクラスタ数k個に分類する非階層型クラスタリングのアルゴリズム
密度準拠型クラスタリング	ある空間に点集合が与えられたとき、互いに密接にきっちり詰まっている点をグループにまとめ、低密度領域にある点(その最近接点が遠すぎる点)を外れ値とする手法
次元の呪い	データの次元が大きくなると、そのデータを分析する際の計算量が指数関数的に増大する現象を指す。 次元の呪いを回避する為、一般的に機械学習の高次元データは次元を減らす。

A-2. Kaggleとは

一般事項	概説
対象層	データサイエンスの開発未経験者
目的	開発技術(データサイエンス)の向上及びその客観的証明
定義	最適予測モデルの形成による法人向けコンサルティングに係るプラットフォーム
主な工程	1. ローデータの確認 2. 前処理 3. 探索的データ解析 4. 特徴量エンジニアリング 5. 予測モデルの形成 6. 最適化
出題傾向	回帰、時系列、分類、予測
形態	1. 企業・官庁等の法人が、Kaggleにデータを共有し、委託する。 2. 各コンペティション参加者が最適予測モデルを競争する。 3. 順位等に応じ、賞金や称号を得ることができる。
開発言語	Python3/R/Julia
権威	1. Google LCC傘下の外資系IT企業主催 2. 開発技術(データサイエンス)の採用基準とする事例あり。

A-3. 補足(Kaggle)

1. Kaggleで上位に入ると、データサイエンスの開発経験に準ずると認識されている。
2. したがって、データサイエンティストの案件に参入しやすくなる可能性がある。
3. 但し、Kaggleで上位に入るには、相当な努力と時間を必要とする。
4. Kaggleの開発言語は、Python3/Rを中心とする。
5. Kaggleの公用語は、質疑応答を含め、すべて英語である。
6. データサイエンティストに就けば、定量分析手法の開発経験及びリポーティング・プレゼンテーション能力こそが最大の技術指標となる。
7. 以上から、Kaggleとは付き合いが肝心である。

A-4. 追加問題(Kaggle)

フェーズ	概要	URL
前提	- Kaggleのアカウントを作成する。 - Kaggleのワークフローを調査する。	Kaggle(HP) 参考(アカウント作成)
開発演習 (Python3/R/Julia)	- 分類問題 - 機械学習 - 前処理 ~ 分類のフローを学ぶ。	Titanic: Machine Learning from Disaster
	- 回帰問題 - 機械学習 - 前処理 ~ 予測のフローを学ぶ。	House Prices: Advanced Regression Techniques

A-5. 追加問題(Kaggle)の解答例

1. 以下のリンクに様々な解答例がある。
2. これを基に、前処理 ~ 機械学習(分類/回帰)のプロセスを理解されたい。
3. 尚、他者のソースコードを読解すると、コーディングを向上することができる。

追加問題	解答例
Titanic: Machine Learning from Disaster	https://www.kaggle.com/c/titanic/notebooks?sortBy=scoreDescending&group=everyone&pageSize=20&competitionId=3136
House Prices: Advanced Regression Techniques	https://www.kaggle.com/c/house-prices-advanced-regression-techniques/notebooks?sortBy=scoreAscending&group=everyone&pageSize=20&competitionId=5407

A-6. 確率論・統計数学(概念レベル)

1. 以下の書籍に記述されているレベルは理解しておくことを推奨する。
2. 個別の分析手法は、その都度、理論書を指定する。

(推薦書籍)

『生き抜くための高校数学: 高校数学の全範囲の基礎が完璧にわかる本』
『まずはこの一冊から 意味がわかる微分・積分』
『まずはこの一冊から 意味がわかる線形代数』
『まずはこの一冊から 意味がわかる統計学』
『まずはこの一冊から 意味がわかる統計解析』
『まずはこの一冊から 意味がわかる多変量解析』
『まずはこの一冊から 意味がわかるベイズ統計学』

A-7. 確率論・統計数学(数式レベル)

確率論・統計数学を更に学びたい受講者は、以下の書籍に取り組むことを推奨する。

(推薦書籍)

『技術者のための基礎解析学 機械学習に必要な数学を本気で学ぶ』

『技術者のための線形代数学 大学の基礎数学を本気で学ぶ』

『技術者のための確率統計学 大学の基礎数学を本気で学ぶ』

A-8. 確率論・統計数学(専門レベル)

これ等の推薦書籍を凌駕する数式及びその応用に興味のある受講者は、以下の書籍に取り組むことを推奨する。

(推薦書籍)

『現代数理統計学の基礎』

『確率微分方程式とその応用』

『これなら分かる最適化数学―基礎原理から計算手法まで』

A-9. 参考(SQL/R/Python)

開発言語の参考資料を下記に提示する。
Google Chromeのブックマーク等を利用して開発時に参照されたい。

- 逆引きSQL構文集
- Python言語リファレンス
- RDocumentation

A-10. プログラミングの要領

1. 具体的に何をどのように処理したいかを設計する。
2. ダイアグラムを描画し、各処理の論理関係を整理する。
3. 実行手順を記述する。
4. 不明点を明確化する。
5. この不明点を検索し、例に即して記述する。
6. エラーメッセージを敢えて出力し、Web検索を行う。
7. 検索情報を基に、デバッグ処理を施す。

(補足)

- すべての開発言語に共通する。
- GitHub等の開発ツールについても同様である。