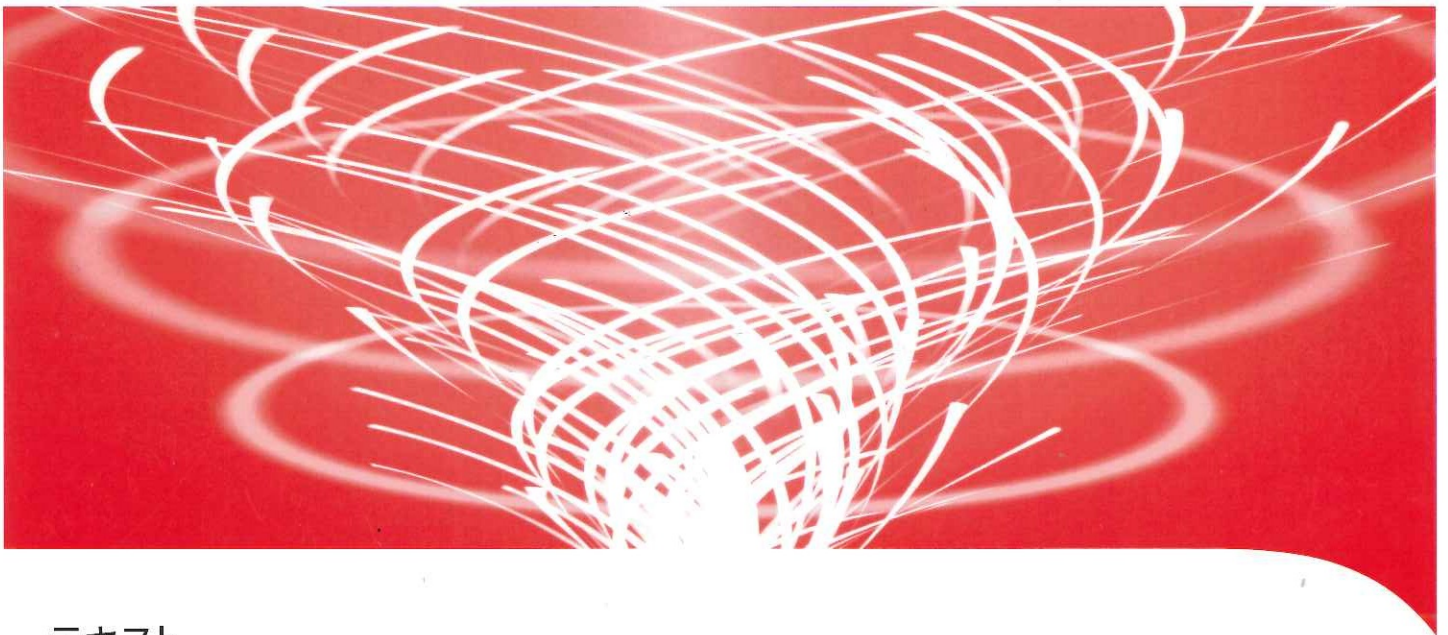


FUJITSU 人材育成・研修サービス

データサイエンス入門

～ データ活用の観点と代表的な分析手法 ～



テキスト

UBU17L1N-01

# 目次

## 第1章 データサイエンスの全体像

1.1 データサイエンスの概要 .....	10
1.1.1 ビジネスにおけるデータサイエンスの活用 .....	10
1.1.2 データサイエンスで実現できること .....	11
1.2 データサイエンスを進めるに当たって .....	12
1.2.1 データサイエンスの実践に必要なもの .....	12
1.2.2 データサイエンスのプロセス .....	13
1.2.3 データサイエンスに必要な人材とスキル .....	14
<参考> スキルチェックリスト .....	15
<参考> データサイエンスプロジェクトの体制例 .....	16
1.3 本コースの学習内容 .....	18

## 第2章 分析手法と評価

2.1 データ分析のアプローチを決める .....	20
2.2 代表的な分析アプローチ .....	21
<参考> 分析の種類と位置づけ .....	22
<参考> 分析事例 .....	23
2.3 分析の前提知識 .....	24
2.3.1 分析ソフトウェア .....	24
2.3.2 データの種類 .....	25
Question データ分析を始める前に...	26
2.4 要約 .....	27
2.4.1 要約とは .....	27
<参考> 要約の観点 .....	28
2.4.2 集計による要約 .....	29
2.4.3 基本統計量による要約 .....	30
(1) 1変数の基本統計量 .....	30
(2) 2変数の基本統計量～相関係数～ .....	32
2.4.4 グラフ化による要約 .....	33
<参考> 量、比率を比較するグラフ .....	34
<参考> 関係性を見るグラフ .....	35
<参考> 分布を見るグラフ .....	36
<参考> 可視化のパターンと、採用するツール .....	37
2.5 クラスタリング .....	38
2.5.1 クラスタリングの位置づけ .....	38
2.5.2 クラスタリングとは .....	39
2.5.3 クラスタリングの手法～k-means法～ .....	40
(1) k-means法とは .....	40
(2) k-means法の流れ .....	41
<参考> クラスタリング結果の活用例 .....	42



<参考> その他のクラスタリングの手法 .....	43
2.5.4 ここまでのまとめ .....	44
Question 仮説検証 .....	45
2.6 統計的検定 .....	46
2.6.1 統計的検定の位置づけ .....	46
2.6.2 統計的検定とは .....	47
2.6.3 統計的検定の用語と考え方 .....	48
2.6.4 統計的検定の手法～t検定～ .....	49
<参考> その他の統計的検定の手法 .....	50
2.6.5 統計的検定のまとめ .....	51
Question 関係性分析 1 .....	52
2.7 回帰分析 .....	53
2.7.1 回帰分析の位置づけ .....	53
2.7.2 回帰分析とは .....	54
2.7.3 回帰分析の手法～線形回帰～ .....	55
(1) 線形回帰とは .....	55
(2) 線形回帰の活用イメージ .....	56
<参考> 線形回帰の種別 .....	57
(3) 線形回帰の流れ .....	58
2.7.4 回帰分析の評価指標 .....	59
<参考> その他の回帰分析の手法 .....	60
2.7.5 回帰分析のまとめ .....	61
Question 関係性分析 2 .....	62
2.8 分類 .....	63
2.8.1 分類の位置づけ .....	63
2.8.2 分類とは .....	64
2.8.3 分類の手法～決定木～ .....	65
(1) 決定木とは .....	65
(2) 決定木の流れ .....	66
2.8.4 分類の評価指標 .....	67
<参考> 交差検証 .....	68
<参考> その他の分類の手法 .....	69
2.8.5 分類のまとめ .....	70
<参考> 分析の種類と位置づけ (再掲) .....	71
<参考> 従来の機械学習とディープラーニングの違い .....	72
2.9 テキストを対象にした分析 .....	73
2.9.1 テキスト分析の流れ .....	73
2.9.2 テキスト分析の適用例～共起関係の分析～ .....	74
2.10 分析アプローチのまとめ .....	75
2.11 分析事例 .....	76
2.11.1 事例 1 .....	76
2.11.2 事例 2 .....	77
第 3 章 データの準備 .....	
3.1 分析に必要なデータの形式 .....	80

3.1.1	データ分析に必要なデータ形式	80
3.1.2	機械学習に必要なデータ形式（従来の機械学習）	81
3.1.3	機械学習に必要なデータ形式（ディープラーニング）	82
3.2	非構造化データから構造化データへの加工	83
3.3	構造化データの加工	84
3.3.1	データ加工の手法	84
3.3.2	集約	85
3.3.3	整形	86
3.3.4	変数作成	87
	<参考> 次元削減 / 次元圧縮	88
Question	分析の青写真を描く	89

## 第4章 データ利活用に向けた計画

4.1	データ利活用に向けた計画	92
4.2	アイデア発想	93
4.2.1	発想の起点	93
4.2.2	データからの発想～データマイニング～	94
	<参考> CRISP-DM	95
4.3	分析計画に役立つフレームワーク	96
	<参考> データ利活用フレームワークの例	97
4.4	分析計画を立てる際の考慮事項	98
4.4.1	課題と活用方針の策定	98
4.4.2	活用方針のパターン	99
	<参考> データ分析のアウトプットと、システム・施策側のアウトプット	100
4.5	データ利活用プロジェクトにおけるマネジメントの観点	101
	<参考> データ利活用プロジェクトの役割分担とスケジュール例	102
4.6	まとめ～ありがちな事象とその回避策～	103
	参考文献	105