



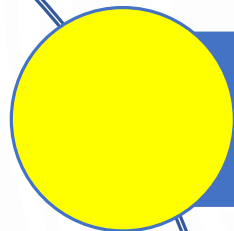
コロナショック時における 日経平均株価とビットコイン価格の予測

法政大学大学院 理工学研究科

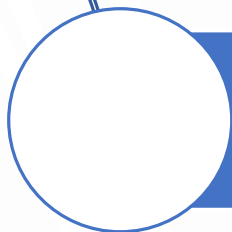
システム理工学専攻

伊藤 翼

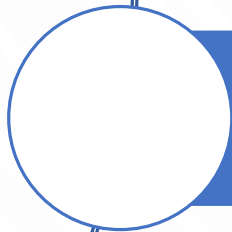
目次



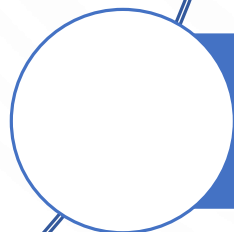
研究背景と目的



分析手法の説明



分析



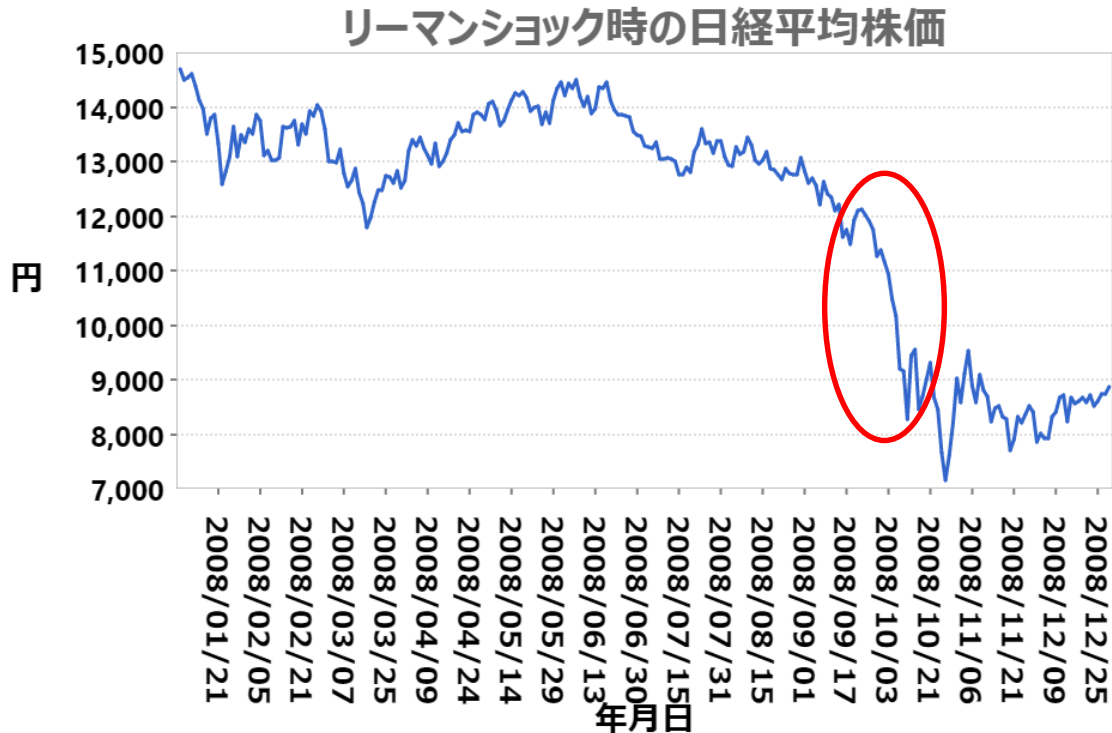
まとめ

研究背景

- 株式や仮想通貨などの「**将来の価格**」を予想することは極めて難しいと言われているが、機関投資家などは市場全体の動向や企業のファンダメンタルズなどを緻密に分析することで、株式などへの投資判断を行っている。
- また、近年では「**機械学習**」などにより「**数理モデル**」を用いた株価の予測手法について盛んに研究が行われている。例えば、福田[4]では、**SCW (Soft Confidence-Weighted Learning)**と呼ばれる学習方法を用いることで、2015年9月1日から2016年2月29日までの「**日経平均株価**」を対象に価格の「上がる」or「下がる」を**約80%の精度**で予測できたことを報告している。さらに、宮崎・松尾[5]では、**CNN (Convolutional Neural Network)**を用いた深層学習により、2016年1月～6月の「**日経平均株価**」を対象に価格の「上がる」or「下がる」を**平均51%以上の精度**で予測できたことを報告している。

研究背景

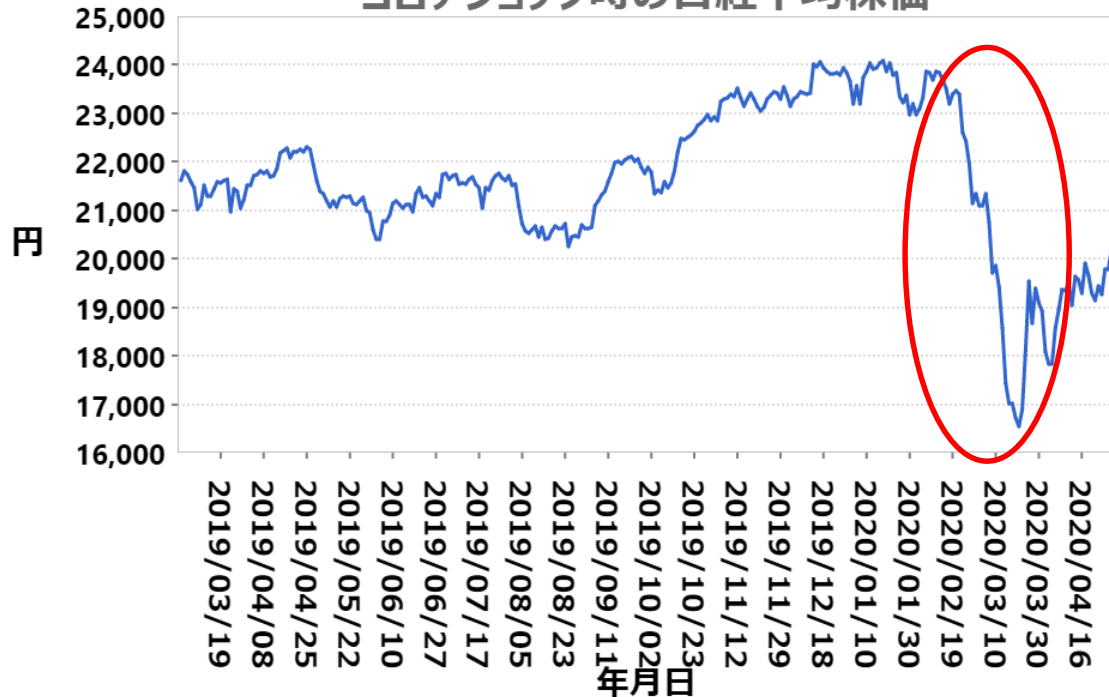
- しかし、このような「数理モデル」を用いた株価の予測は、過去データを用いて予測を行うため、2008年のリーマンショックや2015年のチャイナショックなどの急激な株価の変動に弱いという欠点が指摘されている。



研究目的

- そこで本研究では、「統計的学習」を用いて、コロナショック時（2020年1月～4月）の「日経平均株価」と「ビットコイン価格」の予測を行い、予測精度についての検討を行う。

コロナショック時の日経平均株価



コロナショック時のビットコイン価格



目次

- 研究背景と目的
- 分析手法の説明
- 分析
- まとめ

分析手法を説明するための準備

- 今回の分析では1日間の収益率(Lag1)と2日間の収益率(Lag2)を用いて、価格が「上がる」or「下がる」(Direction)の予測を行う。

表1. Lag1 と Lag2, Direction の例

年月日	株価	Lag1	Lag2	Direction
2020/01/06	23204.86	-0.0076	-0.0112	down
2020/01/07	23575.72	-0.0191	-0.0265	up
2020/01/08	23204.76	0.0160	-0.0034	down
2020/01/09	23739.87	-0.0157	-0.0000	up
2020/01/10	23850.57	0.0231	0.0070	up

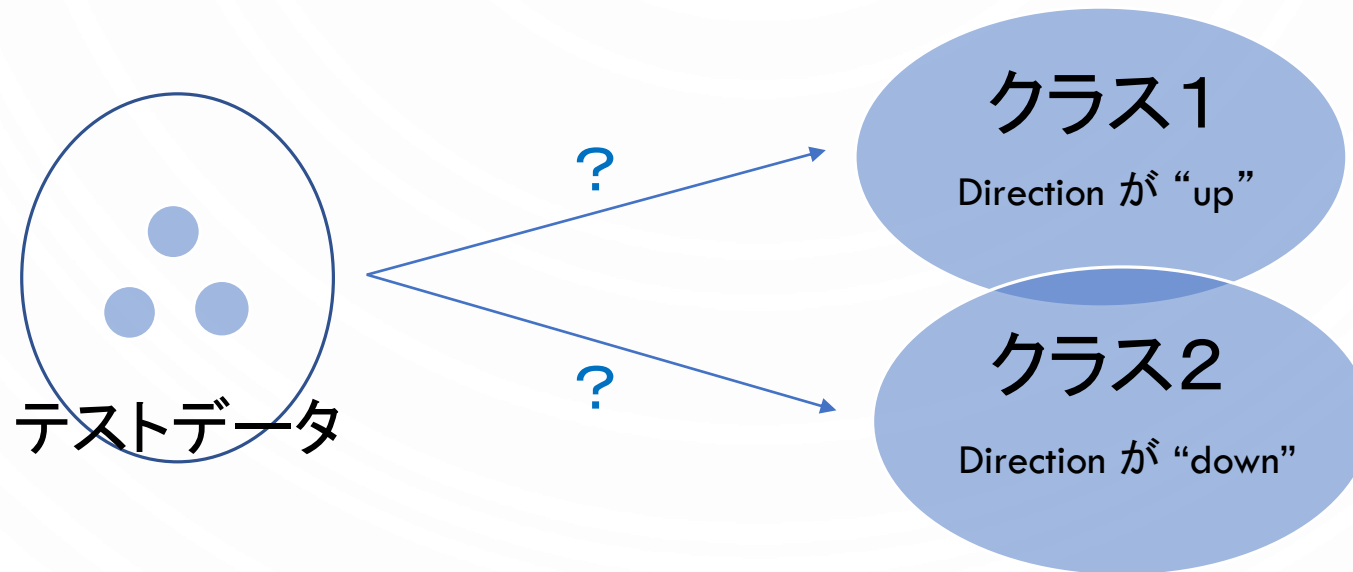
前日から価格が上がったかどうか

$$\frac{23749.87 - 23204.76}{23204.76}$$

$$\frac{23749.87 - 23575.72}{23575.72}$$

分析手法を説明するための準備

- Directionが“up”のデータをクラス1のデータ, Directionが“down”のデータをクラス2のデータと呼ぶことにする.
- 本研究では, テストデータを「クラス1」と「クラス2」に分類することを試みる.



- 分析手法は, ①ロジスティック回帰, ②線形判別分析, ③2次判別分析, ④K最近傍法 の4つを採用する. ([1]を参考にする)

(多重)ロジスティック回帰

X_1 : Lag1, X_2 : Lag2 とし, $X = (X_1, X_2)$ とする.

株価の上昇確率を

$$p(X) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2}}$$

$\beta_0, \beta_1, \beta_2$ は訓練データから最尤推定する.

で表し ($0 < p(X) < 1$), $0 < p(X) \leq 0.5$ の場合は下落 (クラス2), $0.5 < p(X) < 1$ の場合は上昇 (クラス1) に分類する.

(多重)ロジスティック回帰による予測の流れ

訓練データを用いて
パラメータ $\beta_0, \beta_1, \beta_2$ を
最尤推定する。

テストデータを当てはめ、
上昇率 $p(X)$ を計算する。

$0 < p(X) \leq 0.5$
の場合はdown,
 $0.5 < p(X) < 1$
の場合はupと予測。

線形判別分析

X_1 : Lag1, X_2 : Lag2.

クラス1から取ってきたデータ $X = (X_1, X_2)$ は多次元正規分布 $N(\mu_1, \Sigma)$ に従い、
クラス2から取ってきたデータ $X = (X_1, X_2)$ は多次元正規分布 $N(\mu_2, \Sigma)$ に従うとする。

線形判別分析では、予測するデータ $X = x$ を

$\mu_1, \mu_2, \pi_1, \pi_2, \Sigma$ は訓練データから
最尤推定する

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log \pi_k \quad (k = 1, 2)$$

が最大となるクラス k に分類する。

ここで、 π_k はランダムに選んだ観測データがクラス k に属する事前確率。

各クラス間で共通

線形判別分析の流れ

訓練データを用いて
パラメータ $\mu_1, \mu_2, \pi_1, \pi_2, \Sigma$ を
推定する.

テストデータを当てはめ、 $\delta_k(X)$ が
最大となるクラス k に分類する.

注意

- 「ロジスティック回帰」と「線形判別分析」は似た結果を齎すことが知られている。(必ずしもそうとは限らない)
- どちらの手法が優れているかという議論では、「線形判別分析」の場合は各クラスのデータが同じ分散共分散行列を持つ正規分布に従うと仮定しているため、この過程がある程度当てはまる場合は「線形判別分析」の方が優れ、あまり当てはまらない場合は「ロジスティック回帰」の方が優れている。
- P.22以降の分析結果で、似た結果が得られていることを確認して頂きたい。

2次判別分析

線形判別分析との違いは、各クラスで分散共分散行列が異なっているかどうか。

X_1 : Lag1, X_2 : Lag2.

クラス1から取ってきたデータ $X = (X_1, X_2)$ は多次元正規分布 $N(\mu_1, \Sigma_1)$ に従い、

クラス2から取ってきたデータ $X = (X_1, X_2)$ は多次元正規分布 $N(\mu_2, \Sigma_2)$ に従うとする。

2次判別分析では、予測するデータ $X = x$ を

$\mu_1, \mu_2, \pi_1, \pi_2, \Sigma_1, \Sigma_2$ は訓練データから最尤推定する

$$\delta_k(x) = -\frac{1}{2}x^T \Sigma_k^{-1} x + x^T \Sigma_k^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma_k^{-1} \mu_k - \frac{1}{2} \log |\Sigma_k| + \log \pi_k \quad (k = 1, 2)$$

が最大となるクラス k に分類する。

ここで、 π_k はランダムに選んだ観測データがクラス k に属する事前確率。

2次判別分析の流れ

訓練データを用いて
パラメータ $\mu_1, \mu_2, \pi_1, \pi_2, \Sigma_1, \Sigma_2$ を
最尤推定する.

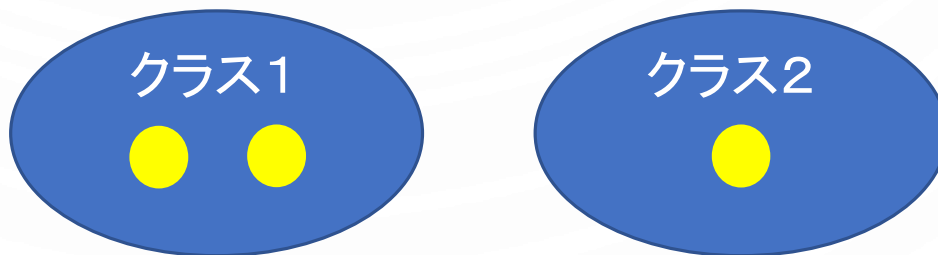
テストデータを当てはめ、 $\delta_k(X)$ が
最大となるクラス k に分類する.

K最近傍法

K : 正の定数, x_0 : テストデータ.

訓練データの中から x_0 に近い K 個のデータを探し, その K 個が属する割合が最も大きいクラスにテストデータを分類する. (本研究では $K = 3$ とする)

ex. $K = 3$ のとき, あるテストデータ x_0 に対してそのデータに近い3つがそれぞれ「クラス1」, 「クラス2」, 「クラス1」に属していたとすると, 「クラス1」に属する割合は $\frac{2}{3}$, 「クラス2」に属する割合は $\frac{1}{3}$ となり, $\frac{2}{3} > \frac{1}{3}$ よりテストデータ x_0 を「クラス1」に分類する.



K最近傍法の流れ

訓練データの中からテストデータに最も近いK個のデータを探す。

そのK個のデータが属するクラスのうち割合が最も高いクラスに分類する。

目次

- 研究背景と目的
- 分析手法の説明
- 分析
- まとめ

分析

- ① 2020年1月～4月の「**日経平均株価**」に対して、訓練データとして直近1年(12ヶ月)、直近2年(24ヶ月)、直近3年(36ヶ月)、直近5年(60ヶ月)、直近10年(120ヶ月)、直近20年(240ヶ月)、直近30年(360ヶ月)の「**日次データ**」を訓練データとして用いた場合で、月ごとに「**各日のDirection**」の予測を行う。
- ② 2020年1月～4月の「**ビットコインの価格**」に対して、訓練データとして直近1年(12ヶ月)、直近2年(24ヶ月)、直近3年(36ヶ月)、直近5年(60ヶ月)の「**日次データ**」を訓練データとして用いた場合で、月ごとに「**各日のDirection**」の予測を行う。

ビットコインは、取れたデータ数の関係から、直近5年分まで分析を行う。

ex. 直近1年のデータを用いて予測を行う場合

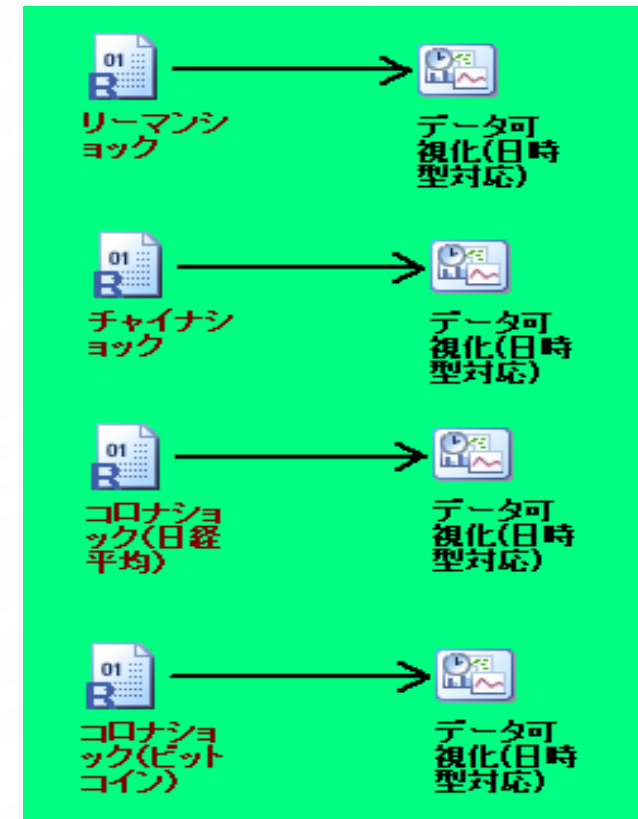
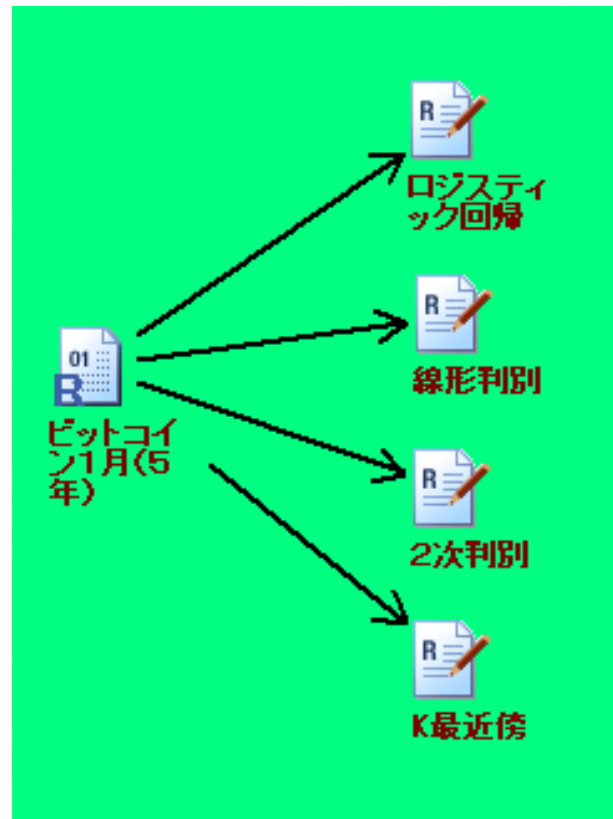
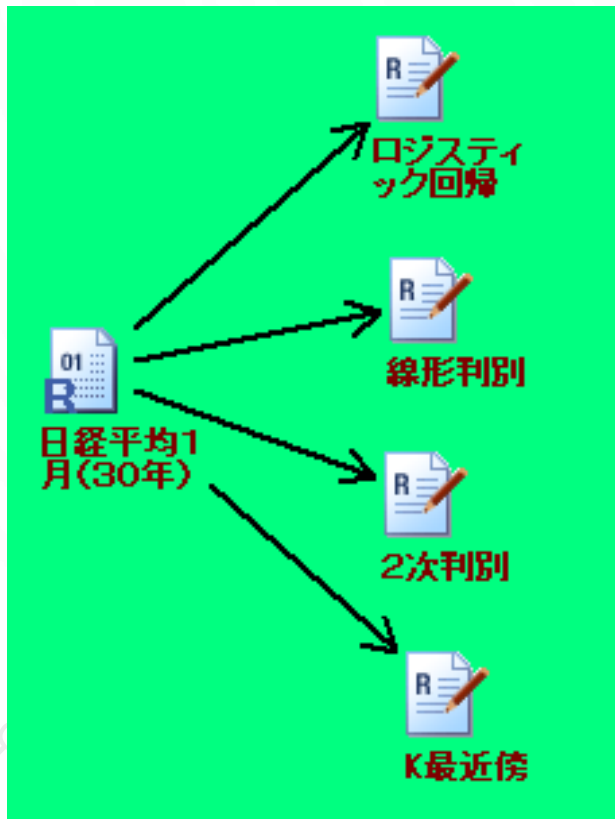
2020年1月の「**各日のDirection**」を予測する時は、直近12ヶ月の2019年1月～同年12月の「**日次データ**」を訓練データとして用いる。

2020年2月の「**各日のDirection**」を予測する時は、直近12ヶ月の2019年2月～2020年1月の「**日次データ**」を訓練データとして用いる。

以下同様に、1ヶ月ずつ訓練データをずらして予測を行う。

VRPの画面

既存のRのみで分析を行おうとすると、分析の度にコマンドを打たなければならないが、VRPの「Rスクリプト」で分析手法を保存しておくことができるのでデータを読み込んで少し操作するだけで分析結果を出力できる。また、分析ごとにファイルが増えていかず、1つの画面上で見やすく分析を行うことができる。さらに、株価のグラフを作成する時には「データ可視化(日時型対応)」で綺麗なグラフを書くことができる。



次ページ以降の分析結果の見方

この箇所には, 分析結果の中で注目して頂きたいポイントを記載している。

直近〇年のデータを用いた予測の正解率

2020年 月	ロジスティック 回帰	線形判別 分析	2次判別 分析	K最近傍法
1月				
2月	この箇所には, 月ごとに各日の 予測の正解率を記載している。			
3月				
4月				



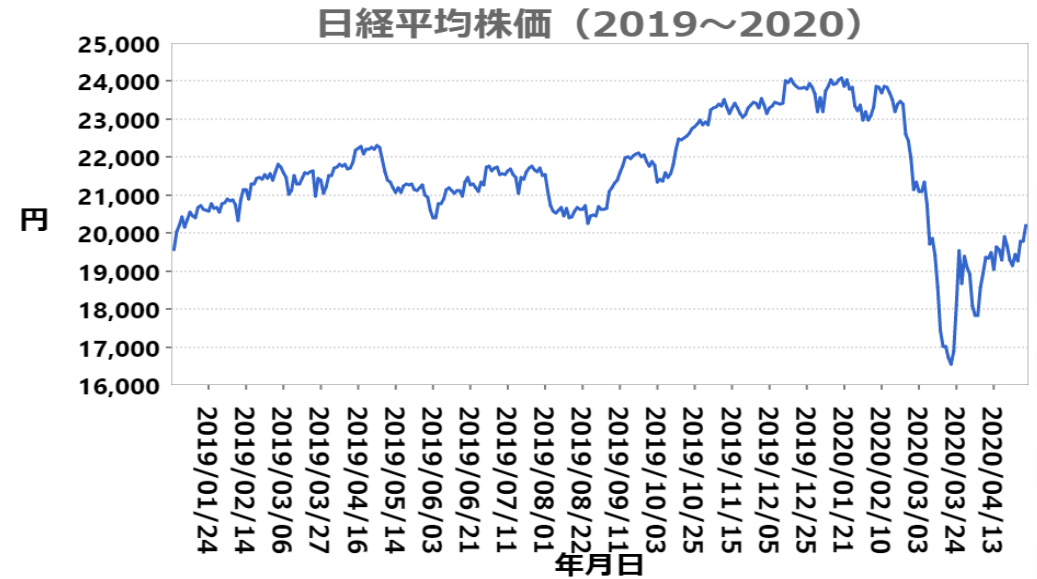
ロジスティック 回帰	線形判別 分析	2次判別 分析	K最近傍法
down up (The 'up' is circled in blue)	down up	down up	down up
down up (The 'up' is circled in blue)	この箇所には, 4か月間を通して, どのように予測をして, 実際はどちらだったのかを記載している。		

分析結果(日経平均株価)

・「K最近傍法」以外の3つの手法については2019年の株価は上昇基調であったことに影響され、“up”と予測しているものが多く、その結果コロナショック時(2020年2月~3月)の正解率が低くなっていることが読み取れる。
 ・「K最近傍法」は、分析対象のトレンドに結果が左右されづらく、コロナショック時にも比較的高い正答率である。

直近1年のデータを用いた予測の正解率

2020年 月	ロジスティック 回帰	線形判別 分析	2次判別 分析	K最近傍法
1月	57.89%	57.89%	47.37%	68.42%
2月	38.89%	38.89%	50.00%	50.00%
3月	47.62%	47.62%	42.86%	66.67%
4月	52.38%	52.38%	61.90%	47.62%



ロジスティック 回帰	down		up		線形判別 分析	down		up		2次判別 分析	down		up		K最近傍法	down		up	
	down	up	down	up		down	up	down	up		down	up	down	up					
down	11	8	10	7	18	14	25	15	25	15	25	15							
up	32	28	33	29	25	22	18	21	18	21	18	21							

分析結果（日経平均株価）

・直近2年分のデータを用いて予測を行った場合は、全ての手法において“up”と予測することが比較的多く、「K最近傍法」以外の3つの手法ではコロナショック時の正解率が低くなっているが、直近1年を用いた場合と比較すると「ロジスティック回帰」と「線形判別分析」では2月の正解率が上がっている。

直近2年のデータを用いた予測の正解率

2020年 月	ロジスティック 回帰	線形判別 分析	2次判別 分析	K最近傍法
1月	57.89%	57.89%	52.63%	57.89%
2月	44.44%	44.44%	38.89%	61.11%
3月	47.62%	47.62%	42.86%	52.38%
4月	57.14%	57.14%	57.14%	23.81%



ロジスティック 回帰	down		up		線形判別 分析	down		up		2次判別 分析	down		up		K最近傍法	down		up	
	down	up	down	up		down	up	down	up		down	up	down	up					
down	11	6	11	6	down	14	12	17	15	down	17	15							
up	32	30	32	30	up	29	24	26	21	up	26	21							

分析結果（日経平均株価）

- ・直近3年のデータを用いて予測を行うと、直近2年分を用いた場合と比べて“up”と予測することが多くなり、その結果「ロジスティック回帰」と「線形判別分析」コロナショック時である2月の正解率が低くなっている。
- ・「K最近傍法」は依然としてショック時にも正解率は安定している。

直近3年のデータを用いた予測の正解率

2020年 月	ロジスティック 回帰	線形判別 分析	2次判別 分析	K最近傍法
1月	57.89%	57.89%	47.37%	52.63%
2月	38.89%	38.89%	44.44%	61.11%
3月	47.62%	47.62%	42.86%	57.14%
4月	52.38%	52.38%	47.62%	33.33%



ロジスティック 回帰	down		up		線形判別 分析	down		up		2次判別 分析	down		up		K最近傍法	down		up	
	down	up	down	up		down	up	down	up		down	up	down	up					
down	6	3	6	3	6	3	8	8	8	14	18	14							
up	37	33	37	33	37	33	35	28	28	22	25	22							

分析結果(日経平均株価)

- ・直近5年分のデータを用いて予測を行うと, 訓練データの上昇傾向がさらに強くなり, 「ロジスティック回帰」と「線形判別分析」の2つは4カ月間通して全て“up”と予測してしまっている。
- ・また, 訓練データとテストデータの傾向が大きく異なるため, 全体的に正解率が低くなってしまっている。

直近5年のデータを用いた予測の正解率

2020年 月	ロジスティック 回帰	線形判別 分析	2次判別 分析	K最近傍法
1月	57.89%	57.89%	47.37%	36.84%
2月	33.33%	33.33%	44.44%	55.56%
3月	42.86%	42.86%	42.86%	33.33%
4月	47.62%	47.62%	38.10%	42.86%



ロジスティック 回帰	down		up		線形判別 分析	down		up		2次判別 分析	down		up		K最近傍法	down		up	
	down	up	down	up		down	up	down	up		down	up	down	up					
down	0	0	0	0	down	0	0	7	9	down	19	22							
up	43	36	43	36	up	43	36	36	27	up	24	14							

分析結果(日経平均株価)

- ・直近10年分のデータを用いて予測を行うと、訓練データの直近5年分を用いた場合と比較して上昇傾向は若干緩和され、正解率は全体的に上昇したように読み取れる。
- ・また、直近10年分を用いることで、初めて「2次判別分析」の3月の正解率が50%を超えた。

直近10年のデータを用いた予測の正解率

2020年 月	ロジスティック 回帰	線形判別 分析	2次判別 分析	K最近傍法
1月	52.63%	52.63%	57.89%	42.11%
2月	38.89%	38.89%	44.44%	61.11%
3月	42.86%	42.86%	57.14%	47.62%
4月	57.14%	57.14%	42.86%	57.14%



ロジスティック 回帰	down		up		線形判別 分析	down		up		2次判別 分析	down		up		K最近傍法	down		up	
	count	count	count	count		count	count	count	count		count	count	count	count		count			
down	6	4	6	4	down	12	8	24	19	down	24	19	17						
up	37	32	37	32	up	31	28	19	17	up	19	17	17						

分析結果(日経平均株価)

・直近20年分のデータを用いて予測を行うと、直近10年分を用いた場合と比較して「2次判別分析」と「K最近傍法」の2月～4月正解率が下がっていて、「ロジスティック回帰」と「線形判別分析」については1月と3月の正解率は変わらないが、4月の正解率は上がっている。

直近20年のデータを用いた予測の正解率

2020年 月	ロジスティック 回帰	線形判別 分析	2次判別 分析	K最近傍法
1月	52.63%	52.63%	63.16%	47.37%
2月	38.89%	38.89%	44.44%	55.56%
3月	42.86%	42.86%	42.86%	47.62%
4月	61.90%	61.90%	38.10%	52.38%



ロジスティック 回帰	down		up		線形判別 分析	down		up		2次判別 分析	down		up		K最近傍法	down		up	
	down	up	down	up		down	up	down	up		down	up	down	up					
down	8	5	8	5	down	14	13	down	19	15	down	19	15						
up	35	31	35	31	up	29	23	up	24	21	up	24	21						

分析結果(日経平均株価)

・直近30年分のデータを用いて予測を行うと、直近20年分のデータを用いた場合と比べて全体的に正解率が上がったように見える。これは、訓練データに下落傾向のものが含まれ、“down”と予測した回数が増えたためであると考えられる。
 ・しかし、直近2年分や3年分のデータを用いた場合の方がコロナショック時の正解率は高い。

直近30年のデータを用いた予測の正解率

2020年 月	ロジスティック 回帰	線形判別 分析	2次判別 分析	K最近傍法
1月	52.63%	52.63%	68.42%	36.84%
2月	44.44%	44.44%	55.56%	61.11%
3月	47.62%	47.62%	42.86%	38.10%
4月	61.90%	61.90%	42.86%	47.62%



ロジスティック 回帰	down		up		線形判別 分析	down		up		2次判別 分析	down		up		K最近傍法	down		up	
	down	up	down	up		down	up	down	up		down	up	down	up					
down	14	9	down	14	9	down	19	14	down	19	19	down	19	19					
up	29	27	up	29	27	up	24	22	up	24	17	up	24	17					

分析結果(ビットコイン)

・直近1年分のデータを用いて予測を行うと, 2019年の上昇傾向が強く, どの方法でも“up”と予測していることが多いように見える. その結果, コロナショック時の正解率が低くなっている傾向が読み取れる.
 ・「ロジスティック回帰」と「線形判別分析」では, 上昇傾向のある4月の正解率がとても高くなっている.

直近1年のデータを用いた予測の正解率

2020年 月	ロジスティック 回帰	線形判別 分析	2次判別 分析	K最近傍法
1月	48.39%	48.39%	58.06%	58.06%
2月	41.38%	41.38%	51.72%	48.28%
3月	48.39%	48.39%	45.16%	61.29%
4月	76.67%	76.67%	56.67%	50.00%



ロジスティック 回帰	down		up		線形判別 分析	down		up		2次判別 分析	down		up		K最近傍法	down		up	
	down	up	down	up		down	up	down	up		down	up	down	up					
down	16	12	16	12	down	18	15	23	18	23	18	23	18						
up	44	49	44	49	up	42	46	37	43	42	46	37	43						

分析結果(ビットコイン)

- ・直近2年分のデータを用いて予測を行うと、直近1年分を用いた場合と比べてコロナショック時の正解率が全体的に落ちているように思える。「2次判別分析」の3月の正解率のみ少しだけ上昇した。
- ・その一方で、「K最近傍法」以外の1月の正解率は上昇している。

直近2年のデータを用いた予測の正解率

2020年 月	ロジスティック 回帰	線形判別 分析	2次判別 分析	K最近傍法
1月	51.61%	51.61%	64.52%	48.39%
2月	41.38%	41.38%	41.38%	41.38%
3月	48.39%	48.39%	48.39%	48.39%
4月	70.00%	70.00%	63.33%	43.33%



ロジスティック 回帰	down		up		線形判別 分析	down		up		2次判別 分析	down		up		K最近傍法	down		up	
	down	up	down	up		down	up	down	up		down	up	down	up					
down	17	14	down	17	14	down	28	23	down	21	27	down	21	27					
up	43	47	up	43	47	up	32	38	up	39	34	up	39	34					

分析結果(ビットコイン)

- ・直近3年分のデータを用いて予測を行うと、「ロジスティック回帰」と「線形判別分析」の2月の正解率はこれまでの中で1番高くなった。しかし、50%を超えることは出来ていない。
- ・また、「K最近傍法」以外の3つの手法の1月と4月の正解率は高くなっている。

直近3年のデータを用いた予測の正解率

2020年 月	ロジスティック 回帰	線形判別 分析	2次判別 分析	K最近傍法
1月	64.52%	64.52%	64.52%	38.71%
2月	44.83%	44.83%	48.28%	31.03%
3月	48.39%	48.39%	45.16%	48.39%
4月	60.00%	60.00%	53.33%	36.67%



ロジスティック 回帰	down		up	線形判別 分析	down		up	2次判別 分析	down		up	K最近傍法	down		up
	down	up	down		up	down	up		down	up	down		up		
down	7	2		down	7	2		down	6	3		down	14	28	
up	53	59		up	53	59		up	54	58		up	46	33	

分析結果(ビットコイン)

- ・直近5年分のデータを用いて予測を行うと、「ロジスティック回帰」と「線形判別分析」の2月の正解率はこれまでの中で一番高くなった一方で、3月の正解率は直近3年と比べると落ちてしまった。
- ・また、1月と4月の正解率はどの方法でも50%を超えていることが読み取れる。

直近5年のデータを用いた予測の正解率

2020年 月	ロジスティック 回帰	線形判別 分析	2次判別 分析	K最近傍法
1月	58.06%	58.06%	64.52%	51.61%
2月	48.28%	48.28%	41.38%	37.93%
3月	45.16%	45.16%	41.94%	54.84%
4月	63.33%	63.33%	53.33%	56.67%



ロジスティック 回帰	down		up		線形判別 分析	down		up		2次判別 分析	down		up		K最近傍法	down		up	
	down	up	down	up		down	up	down	up		down	up	down	up					
down	7	3	down	7	3	down	2	2	down	22	22	down	22	22					
up	53	58	up	53	58	up	58	59	up	38	39	up	38	39					

目次

- 研究背景と目的
- 分析手法の説明
- 分析
- まとめ

まとめ

- 本研究では, コロナショックに焦点を当て, 「ロジスティック回帰」, 「線形判別分析」, 「2次判別分析」, 「k最近傍法」の4つの手法を用いて「日経平均株価およびビットコイン価格」の予測を試みた.
- 分析結果から, 「k最近傍法」以外の3つの手法については, 想定外の事象(2020年2月, 3月)が起こってしまうと予測精度が大きく悪くなることが読み取ることができた.
- その一方で, 「k最近傍法」は想定外の事象が起こった場合でも, 訓練データ次第ではある程度安定して高い予測精度を保つことが分かった.
- また, どの方法においても, 「どのデータを訓練データとして用いるか」によって予測精度が変わることも読み取ることができた.

まとめ

- 以上のことから、「数理モデル」を用いた株価予想においては、市場の動向を正しく把握し、それぞれの場合に対して適切なモデルを選択し、適切な訓練データを用いることが重要であることを改めて確認することができた。



- 今回は、「統計的学習」を用いた4つの手法を用いた場合の予測精度の検討を行ったが、今後の課題として、ニューラルネットワーク等を用いた「機械学習」の手法などを用いた場合の予測精度とも比較をすることで、より研究を深めることができると考えている。

参考文献

[1] 落海浩・首藤信通,「Rによる統計的学習入門」, 朝倉書店, 2018.

[2] 日経NEEDS-FinancialQUEST

<https://finquest.nikkeidb.or.jp/ver2/online/>

[3] Investing.com

<https://www.investing.com/>

[4] 福田ムフタル,「SCWを用いた株価変動予測」, 人工知能学会研究会資料
SUG-FIN-016-02

[5] 宮崎邦洋, 松尾豊,「深層学習を用いた株価予測の分析」, The 31st Annual
Conference of the Japanese Society for Artificial Intelligence, 2017