データ解析

第15回: 実世界

渡辺澄夫

実世界
との戦い

実世界
観測データ
評価法
回帰・判別分析
推定と検定
解析方法
統計的検定
ベイズ法
・階層ベイズ法
時系列予測
因子・主成分
・クラスタ分析

データ解析の例

実世界の例

日本の県庁所在値のデータを使います。政府統計の総合窓口
http://www.e-stat.go.jp/SG1/estat/eStatTopPortal.do

	①世帯数 ②集計世帯数 ③世帯人員 ⑧女性有業数 14米 15パン 27生鮮野菜 61男子洋服 79自動車 87書籍
札幌市 青森市 〇〇町 〇〇村	47市×96次元　の整数

都市の順

札幌市
青森市
盛岡市
仙台市
秋田市
山形市
福島市
水戸市
宇都宮市
前橋市

さいたま市
千葉市
東京都区部
横浜市
新潟市
富山市
金沢市
福井市
甲府市
長野市

岐阜市
静岡市
名古屋市
津市
大津市
京都市
大阪市
神戸市
奈良市
和歌山市

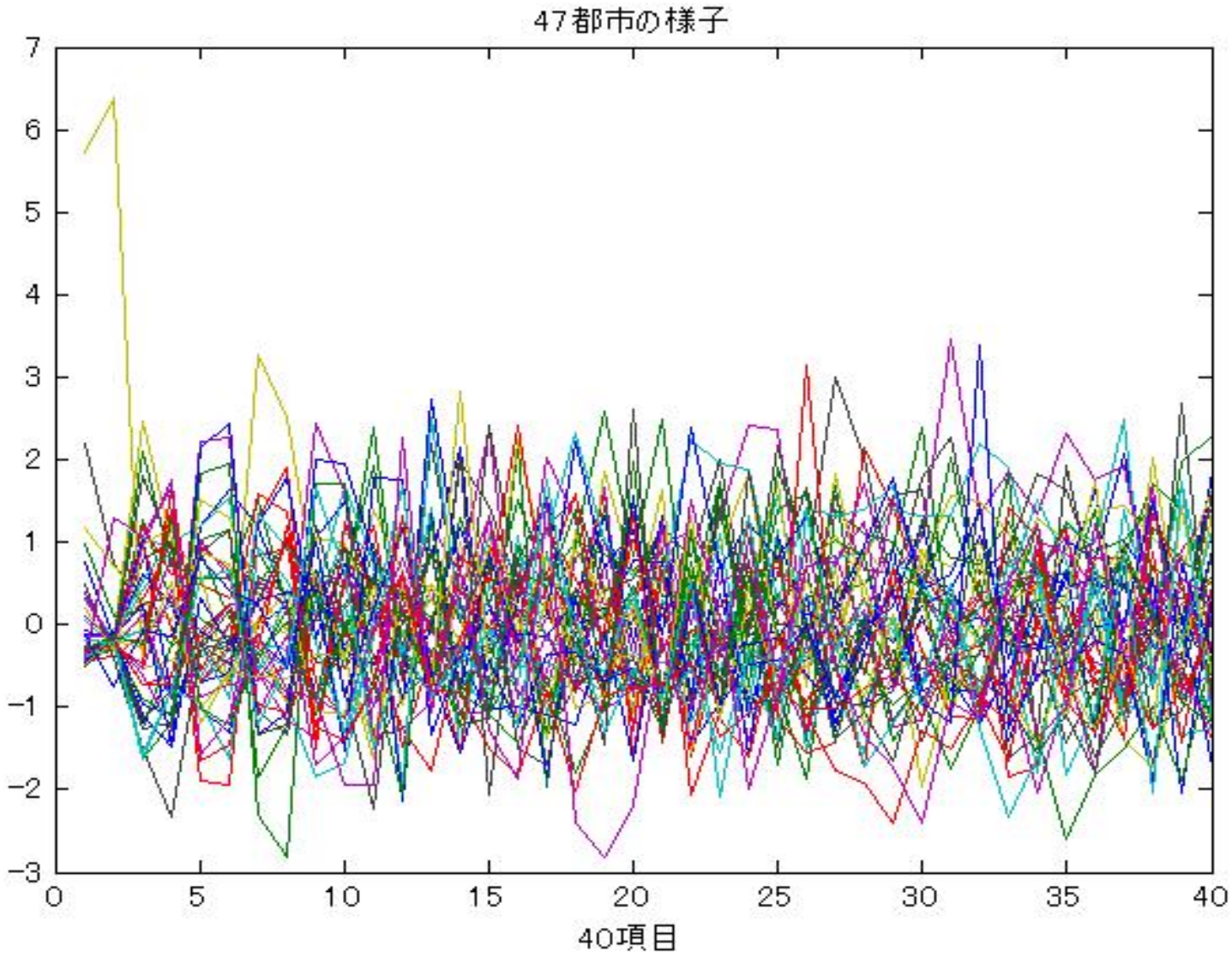
鳥取市
松江市
岡山市
広島市
山口市
徳島市
高松市
松山市
高知市
福岡市

佐賀市
長崎市
熊本市
大分市
宮崎市
鹿児島市
那覇市

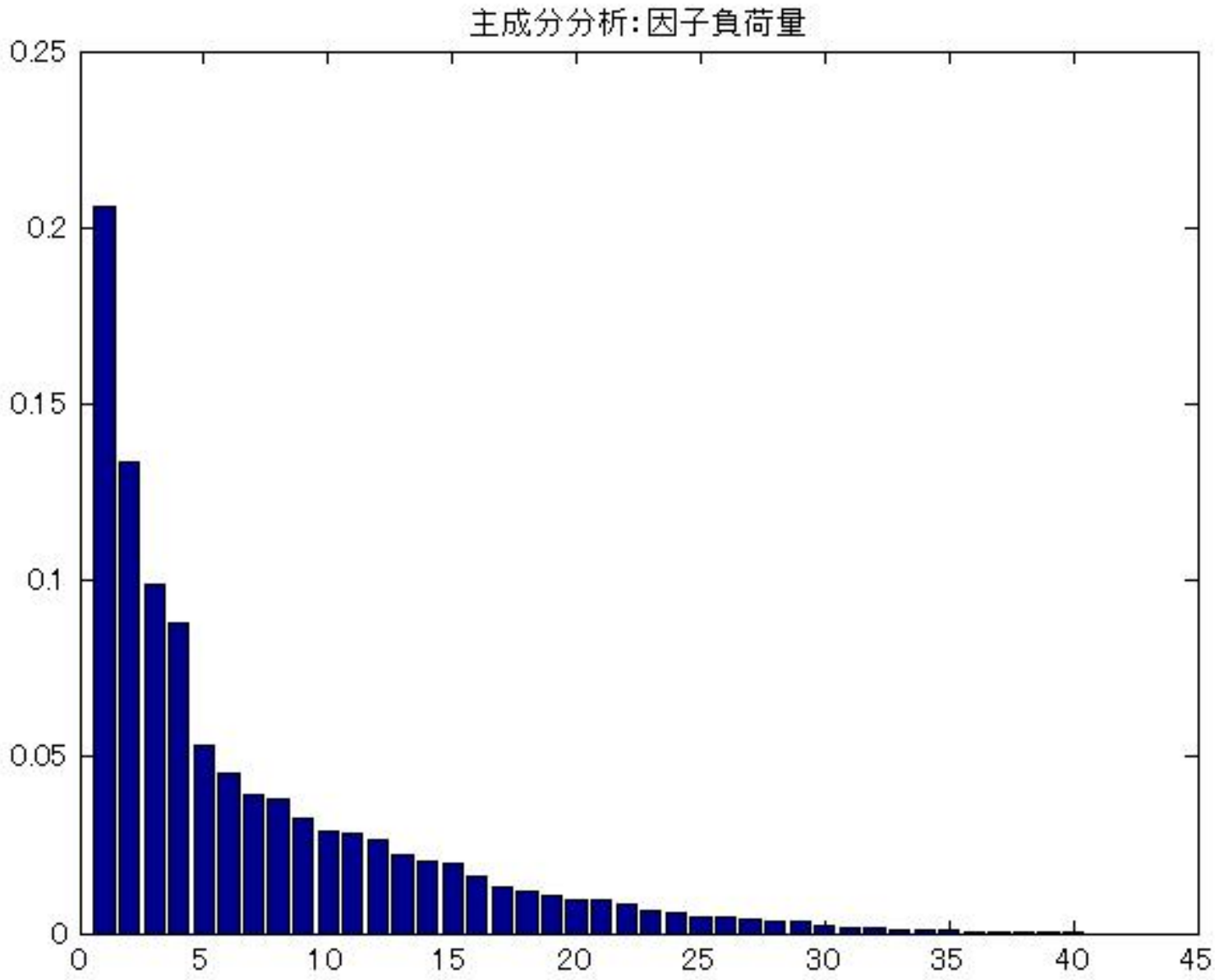
最初の40次元の項目

世帯数割合	
集計世帯数	他の魚介加工品
世帯人員	生鮮肉
18歳未満人員	加工肉
65歳以上人員	牛乳
無職者人員	乳製品
有業人員	卵
世帯主の配偶者女性有業率(%)	生鮮野菜
世帯主の年齢(歳)	乾物・海藻
持家率(%)	大豆加工品
平均畳数	他の野菜・海藻加工品
家賃・地代を支払う世帯	生鮮果物
平均畳数	果物加工品
米	油脂
パン	調味料
麺類	菓子類
他の穀類	調理食品
生鮮魚介	茶類
塩干魚介	コーヒー・ココア
魚肉練製品	茶コーヒーの他の飲料
	酒類

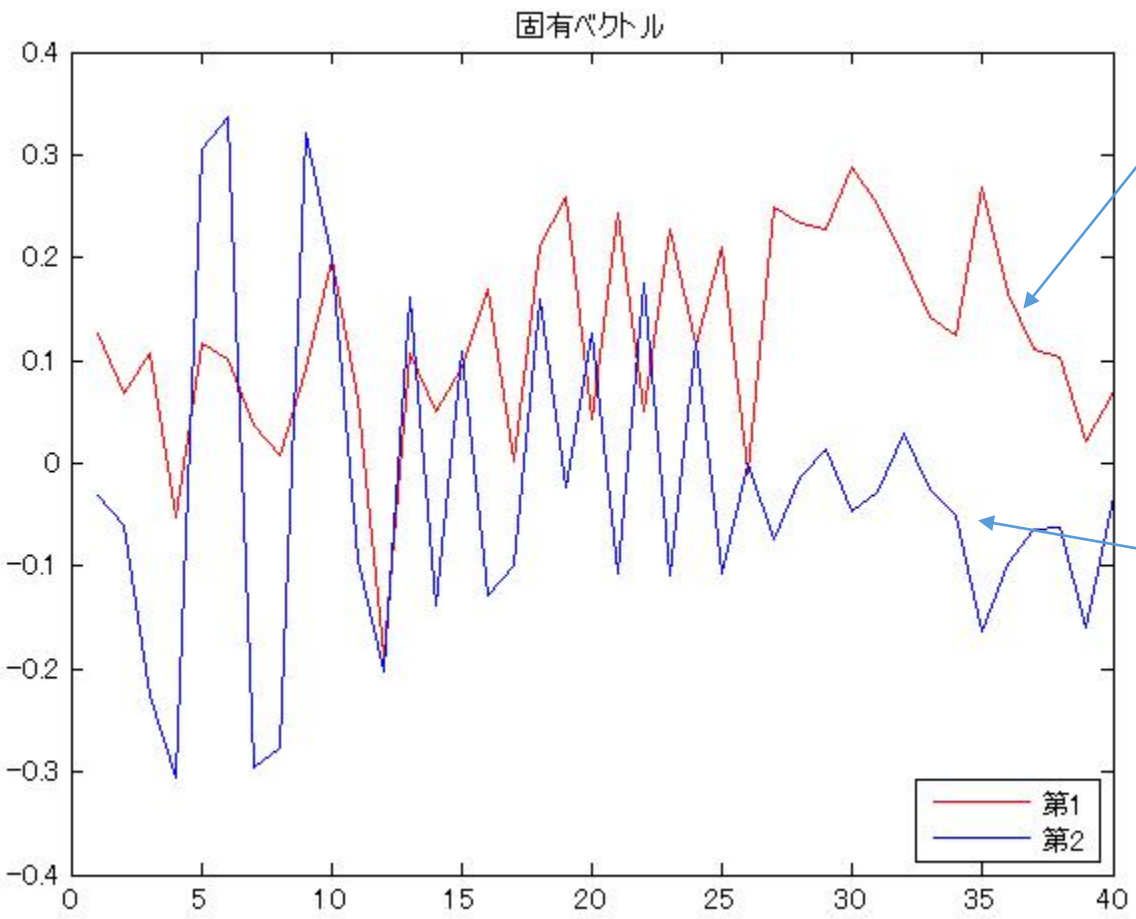
40項目について47都市をプロットした



主成分分析をしてみた　：　因子負荷量



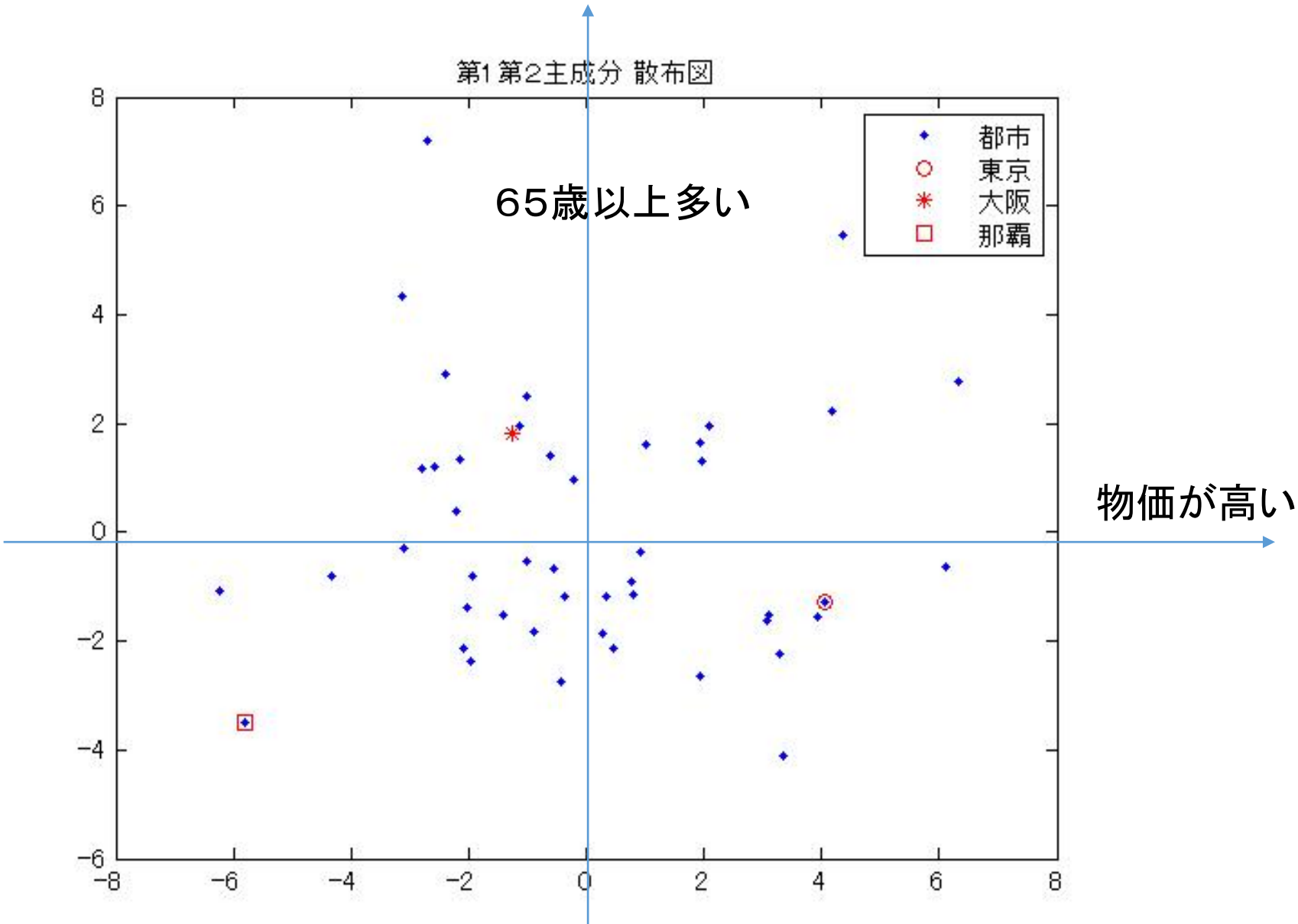
主成分分析をしてみた　：　固有ベクトル



食料にお金が
かかる　物価が高い
加工食品が多い

食料にお金が
かからない
65歳以上が多い
無職が多い
女性有業率低い

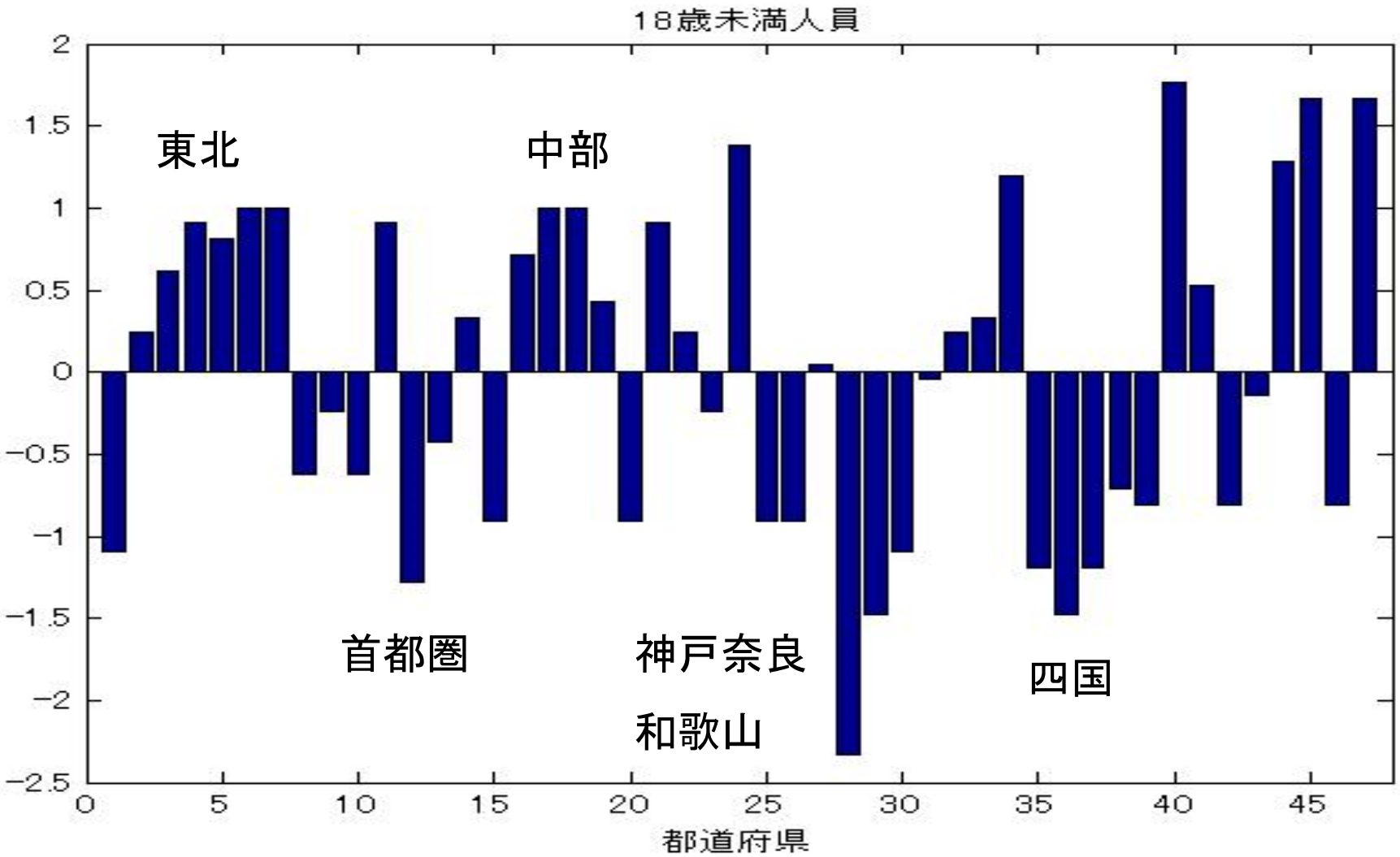
主成分分析をしてみた　：　固有ベクトル



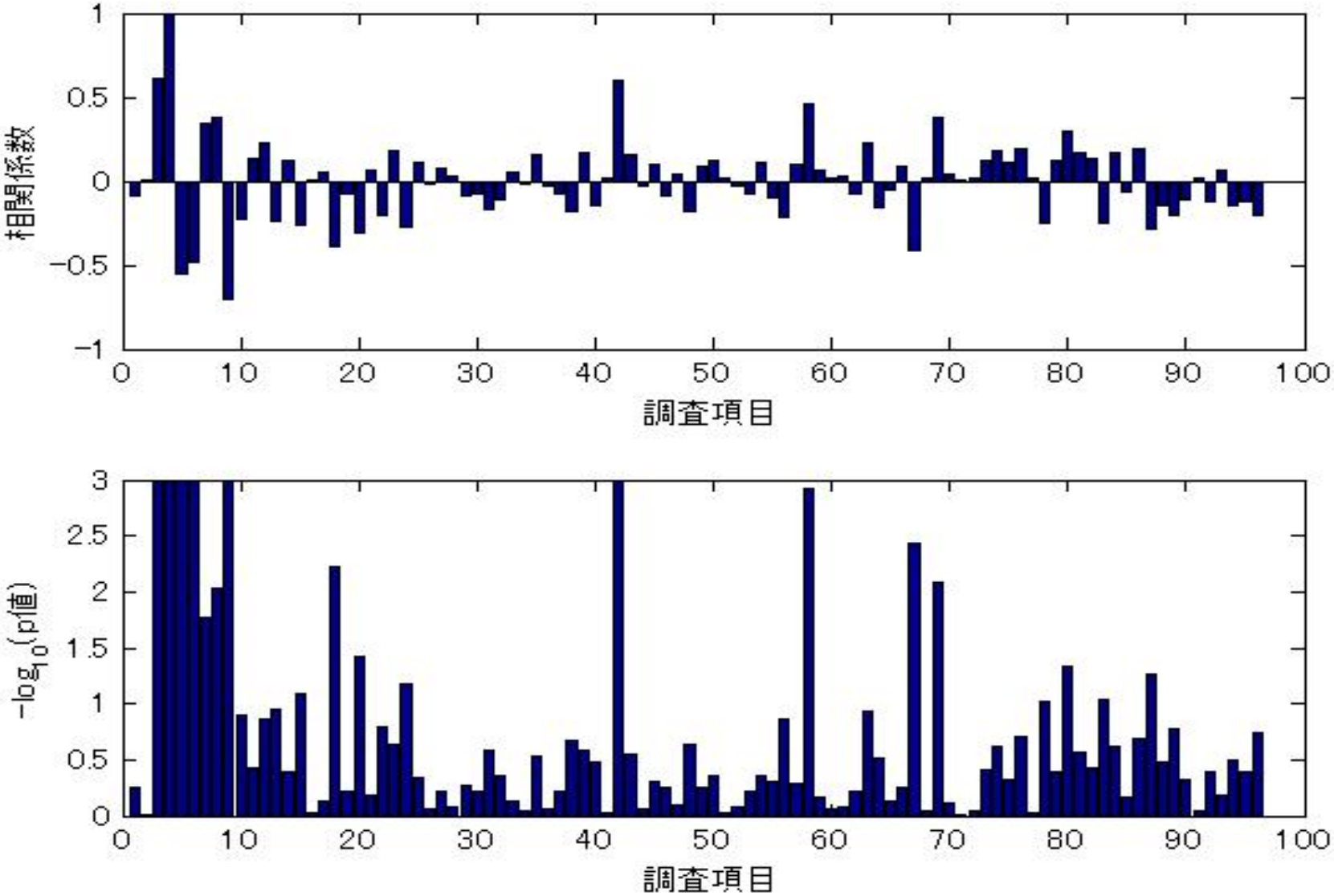
問1

主成分分析によってわかったことを述べよ。

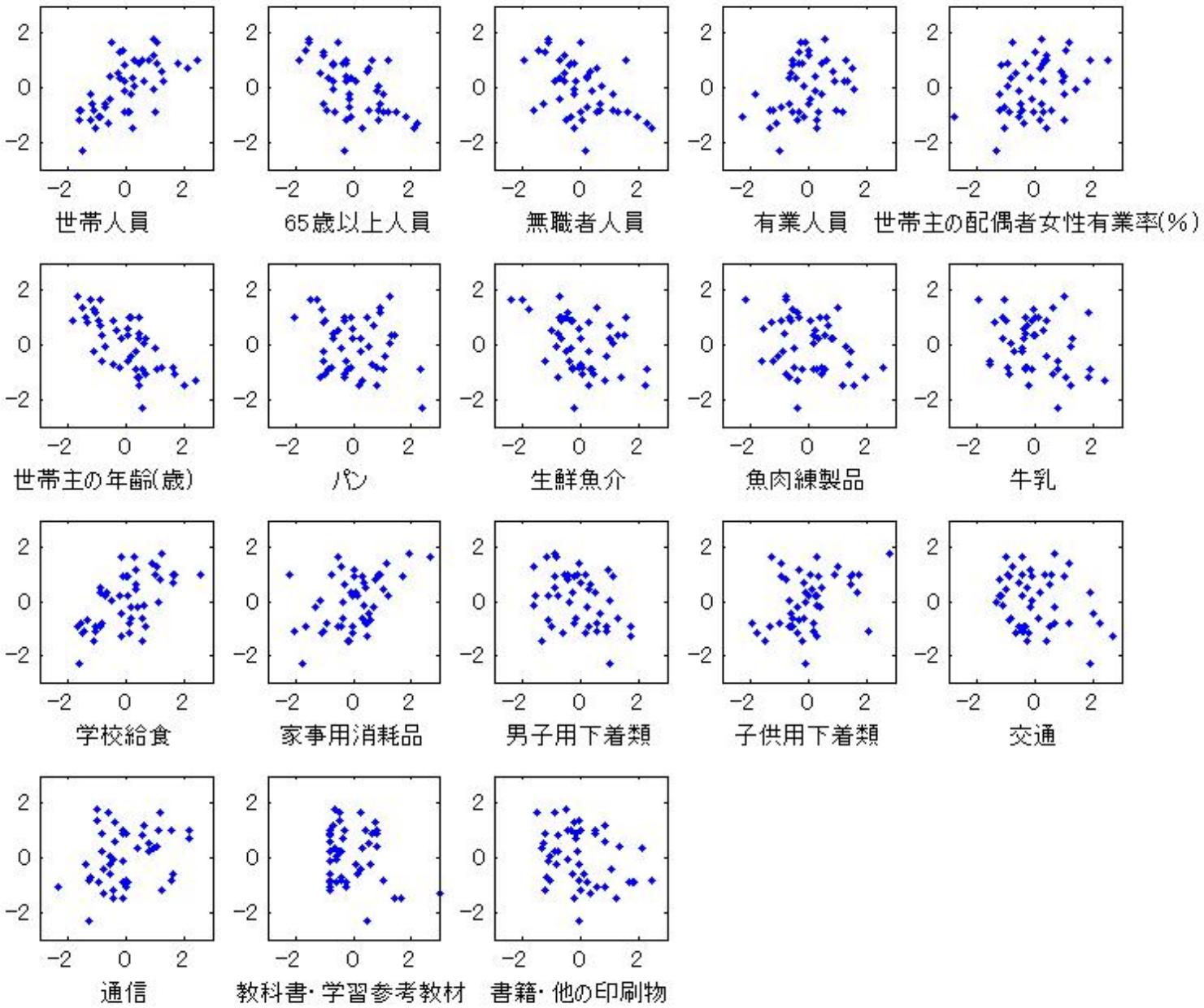
18歳未満の人員について調べてみた



18歳未満の人員と各項目の相関を求めた



相関の大きな項目(p値<0.1)を取り出してプロットした。



線形回帰モデルのAICとBICを計算した。

線形回帰モデル (N=1,2,…,18)

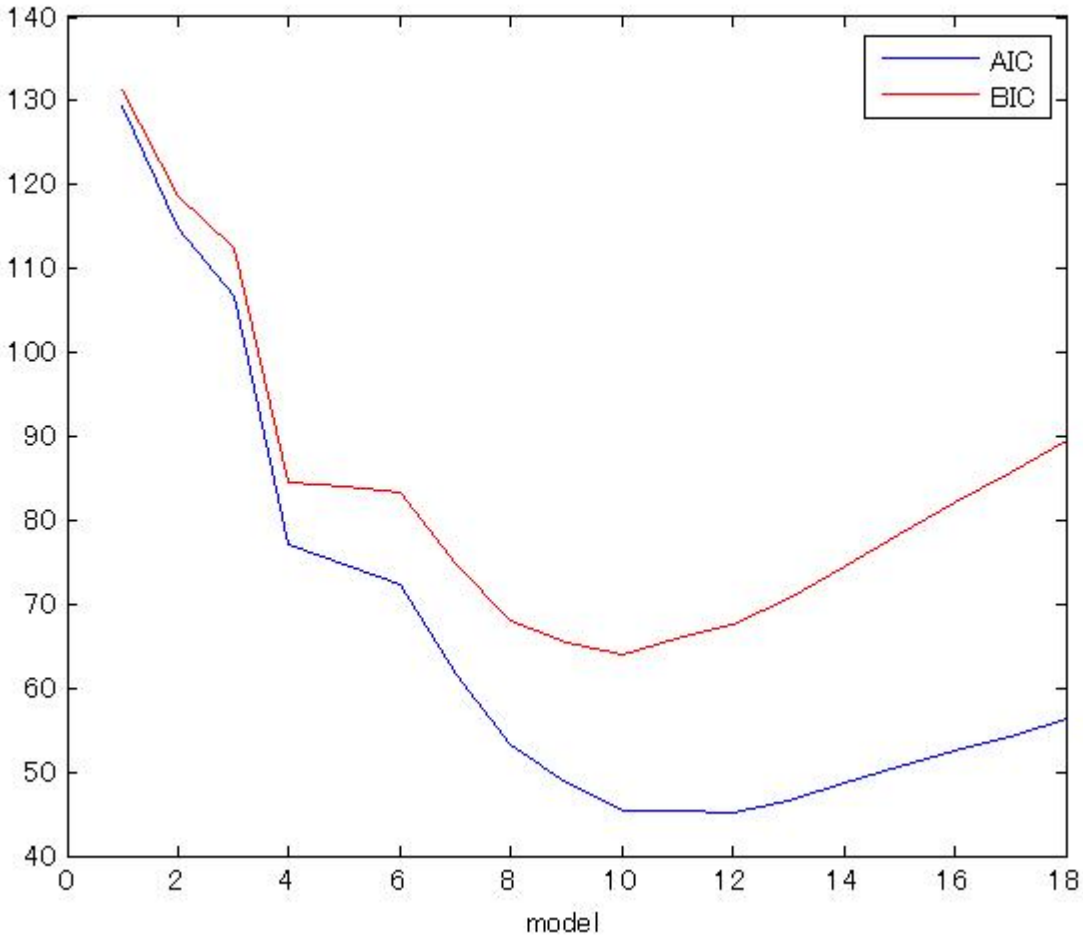
$$Y=a_1x_1+a_2x_2+\cdots+a_NX_N+\text{正規雑音}$$

すなわち統計モデル

$$p(y|x,a,s)= 1/(2\pi s^2)\ \exp(-(1/2s^2)(y-a\cdot x)^2)$$

を用いて AIC と BIC を求めた。

線形回帰モデルのAICとBICを計算した。

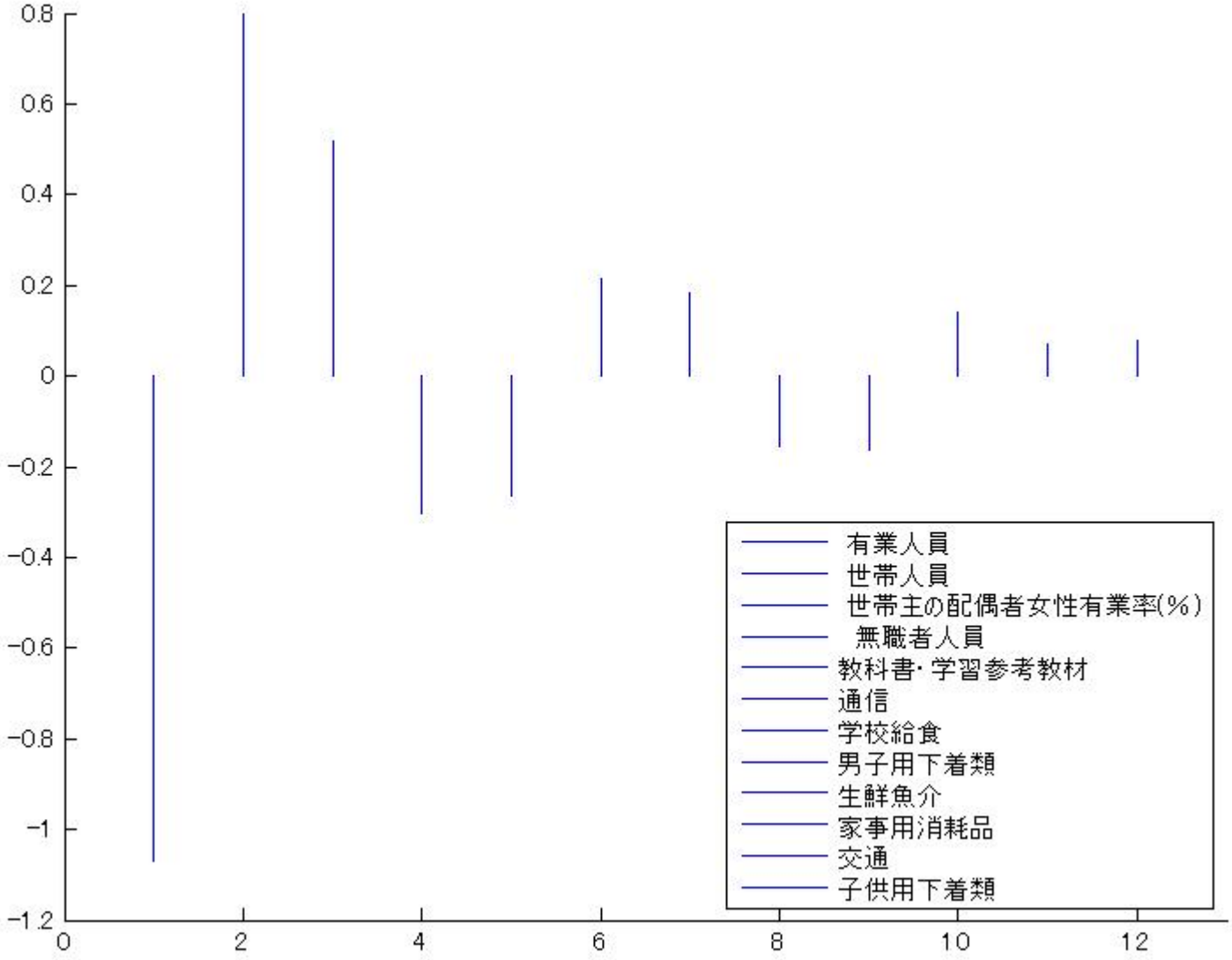


今回は N=12 の

モデルを使う

ことにした。

選ばれたモデルで回帰係数を求めた。



問2

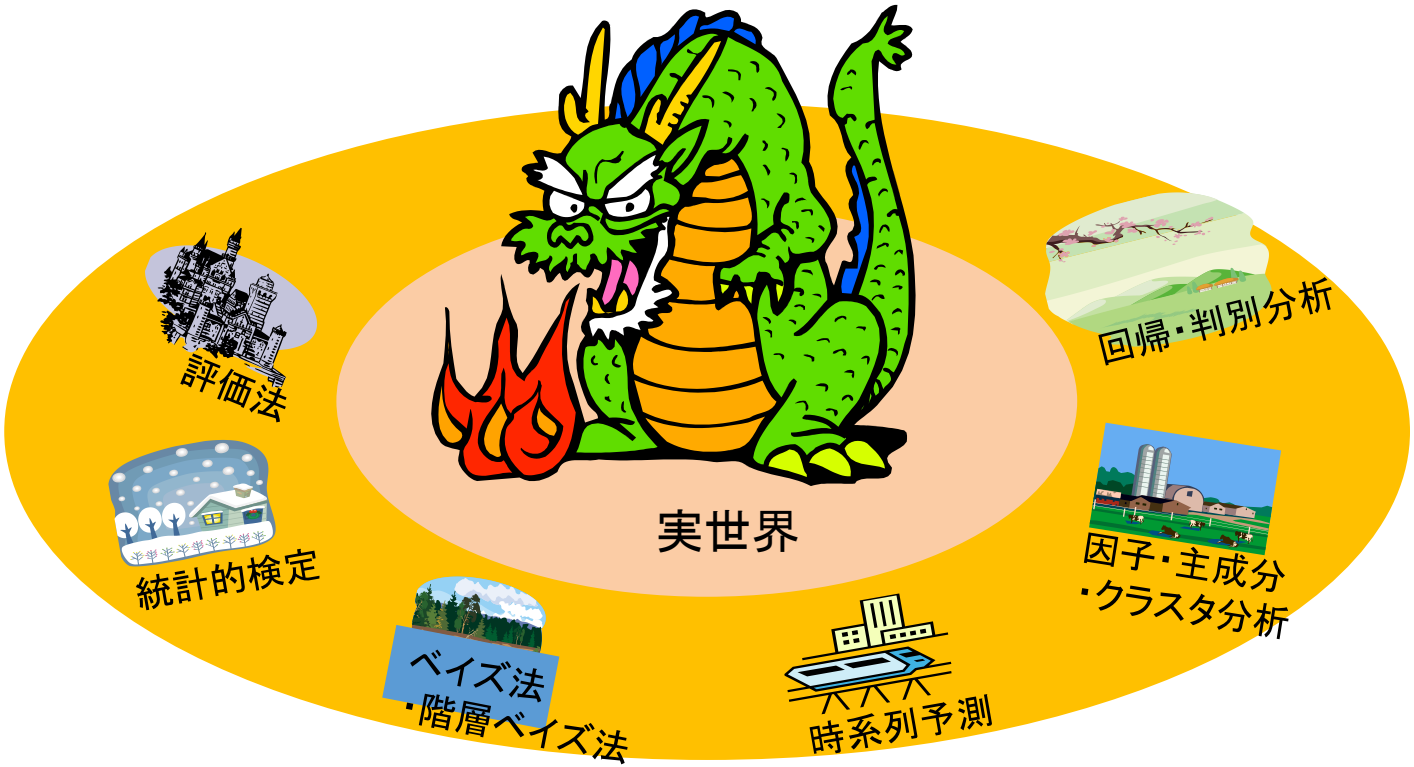
１８歳未満の人員数に関する回帰問題を解くことでわかったことは何か

データ解析の現在

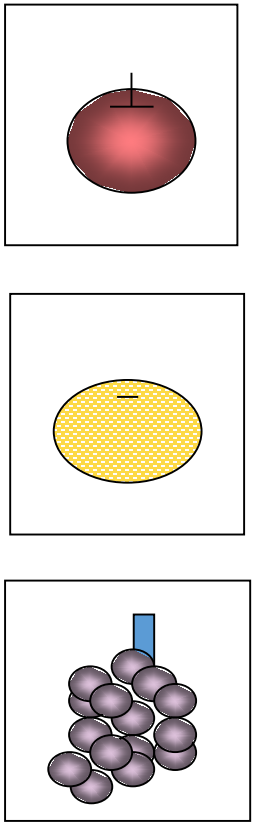
圧倒的に強大な実世界

実世界は、超高次元、際限なく複雑、非定常・・・

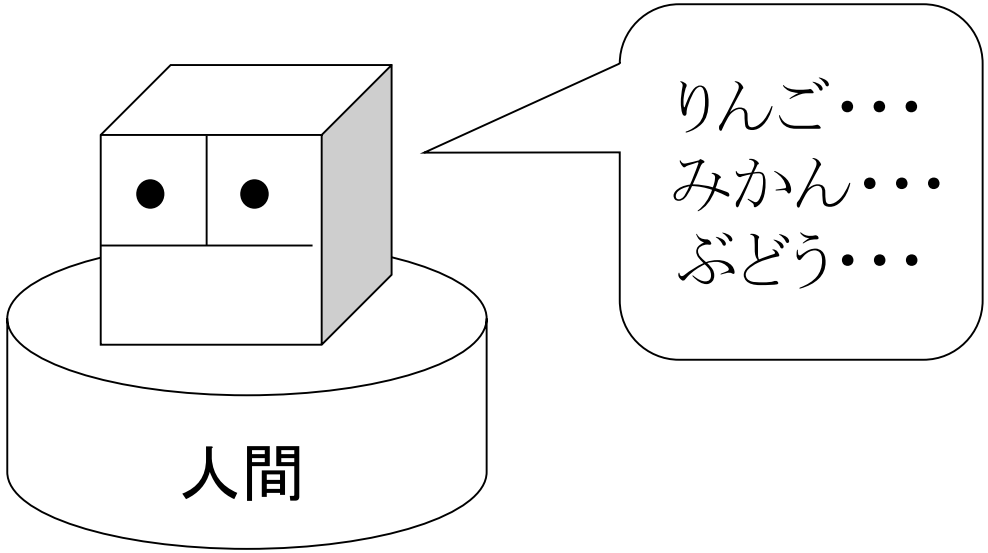
簡単なデータ解析だけでは、もう戦えない・・・



機械学習の登場

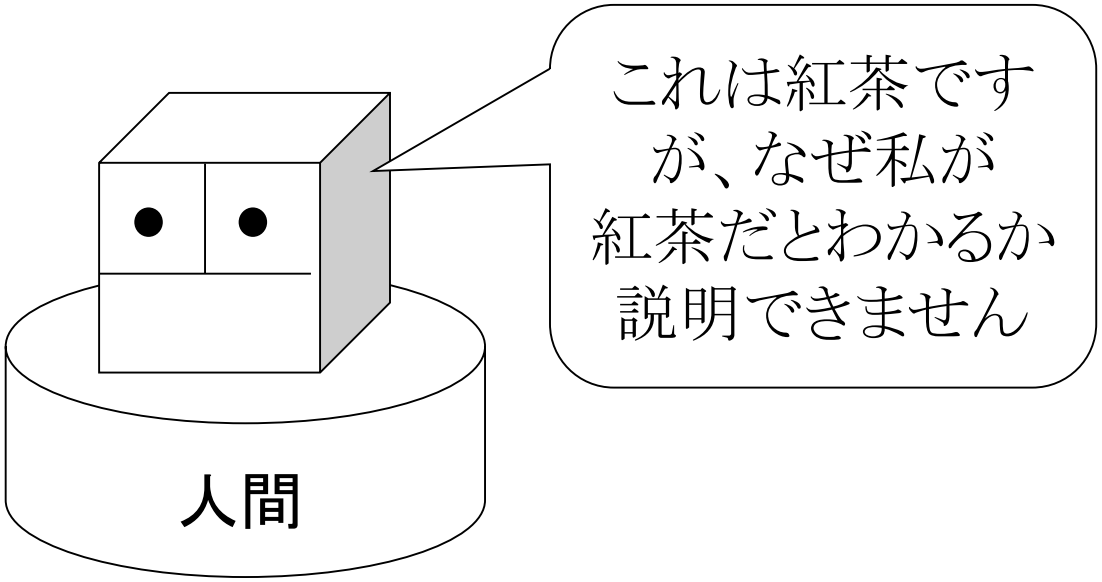


$$y = f(x)$$



人間の関数は

言葉で説明できない



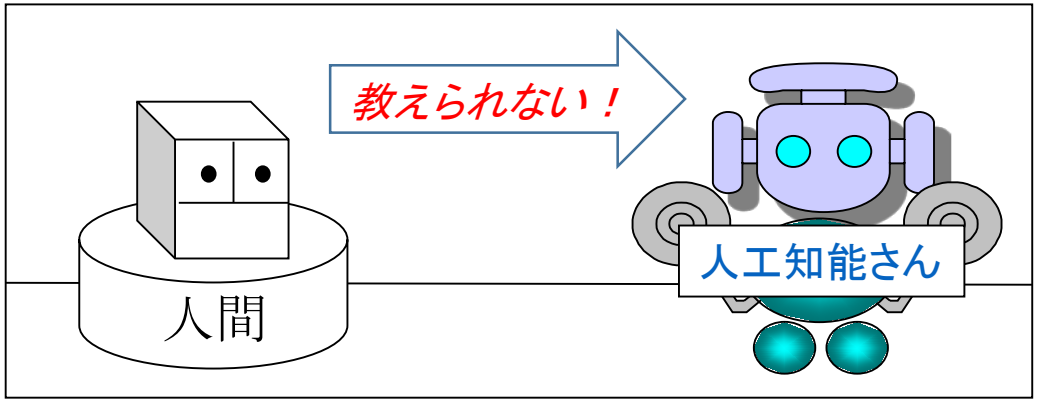
できるけど　説明できない・・・

できるけど　説明できないことは　たくさん　ある

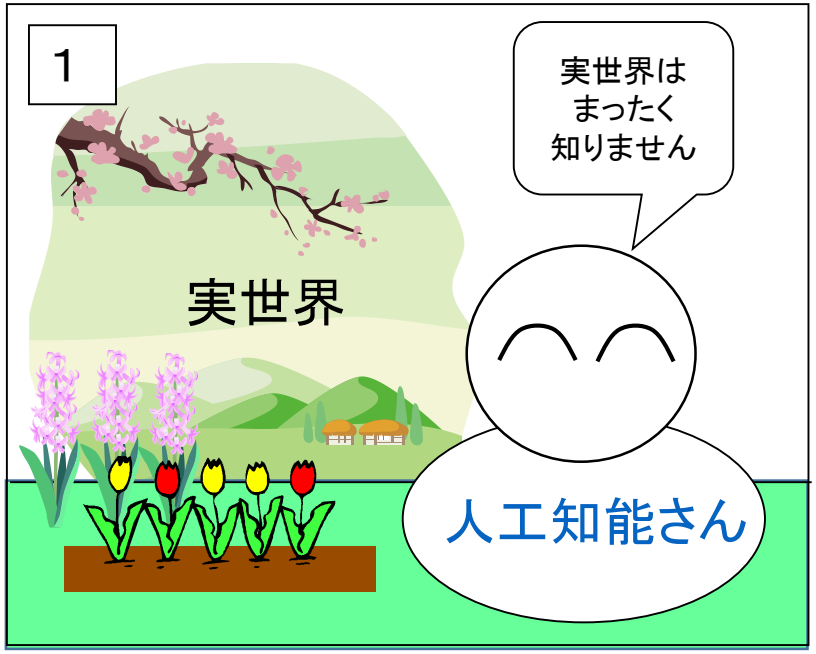
◎　歩く　走る　泳ぐ　自転車に乗る

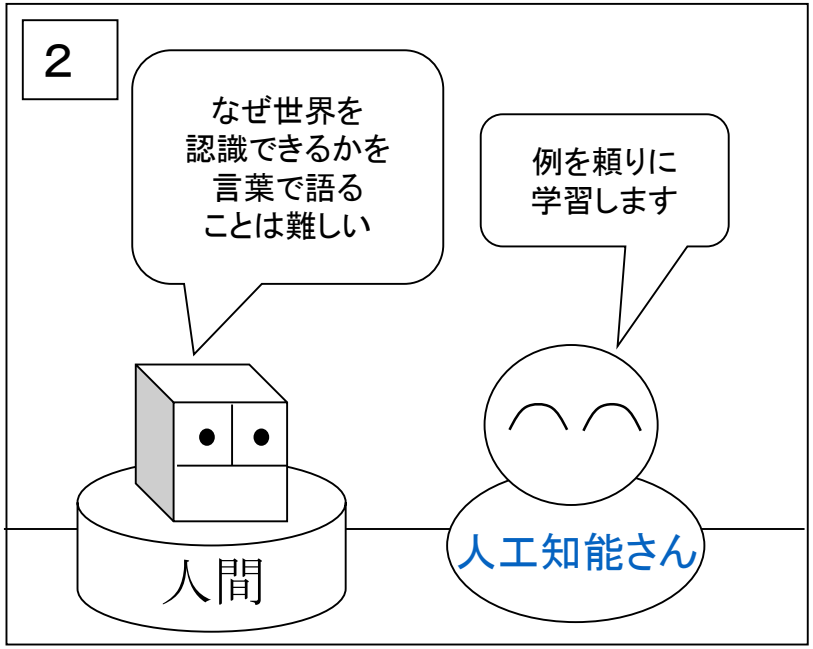
◎　文字を読む　漫画を読む　友達を見つける

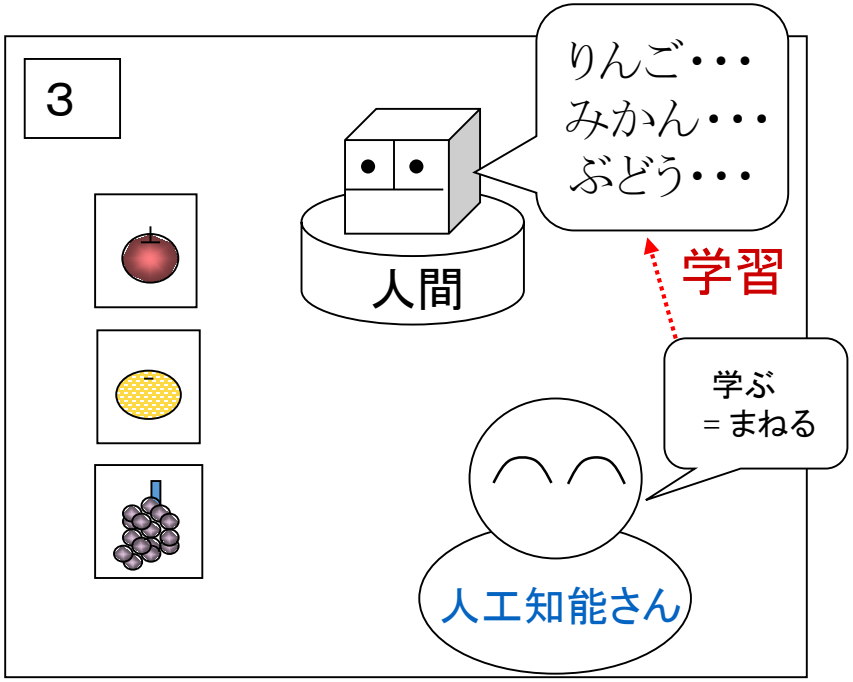
◎　言葉をきく　話しをする

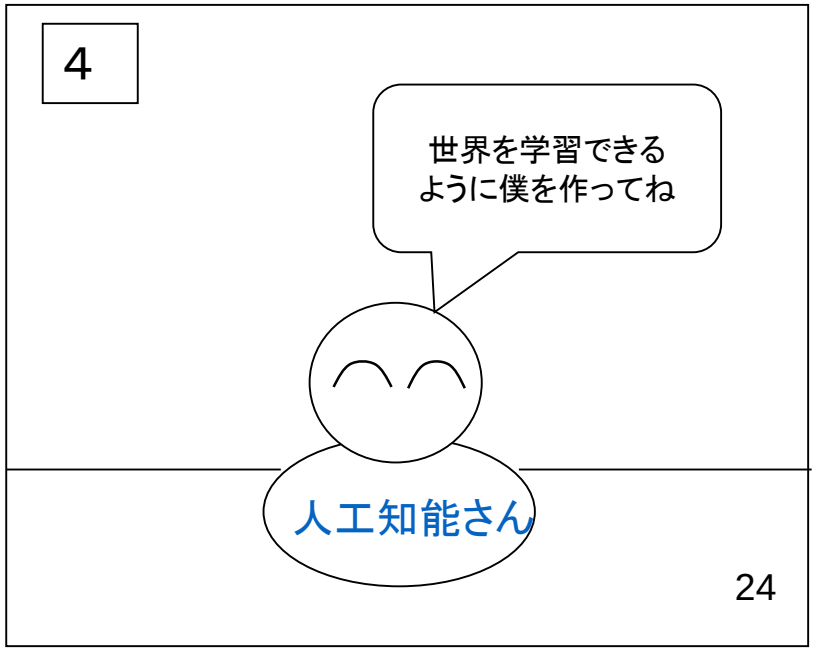


説明できない → 教えられなかった


1
実世界
実世界は
まったく
知りません
人工知能さん

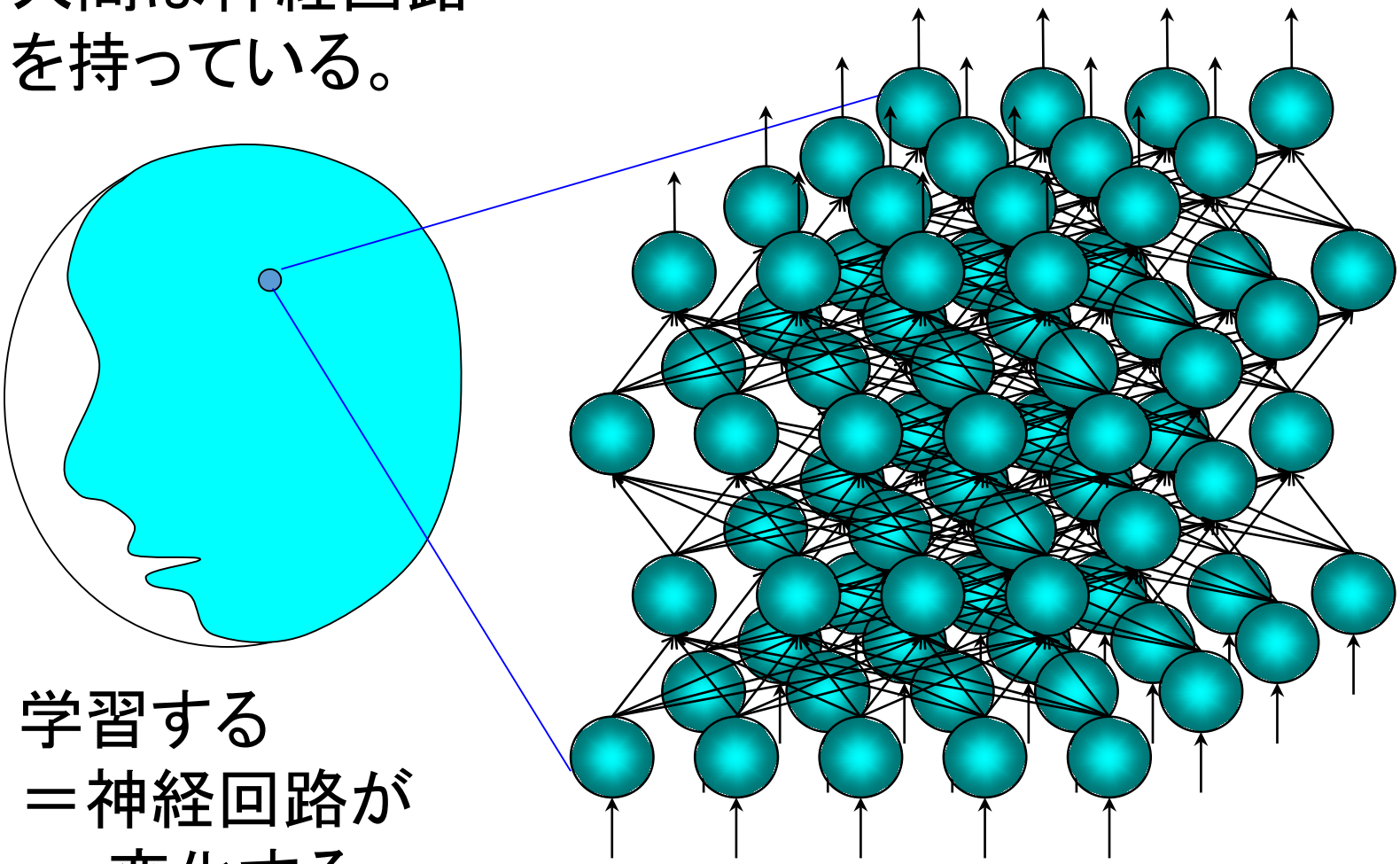

2
なぜ世界を
認識できるかを
言葉で語る
ことは難しい
例を頼りに
学習します
人間
人工知能さん


3
りんご・・・
みかん・・・
ぶどう・・・
人間
学習
学ぶ
＝まねる
人工知能さん


4
世界を学習できる
ように僕を作ってね
人工知能さん

人間や生物が学習できる仕組み

人間は神経回路
を持っている。



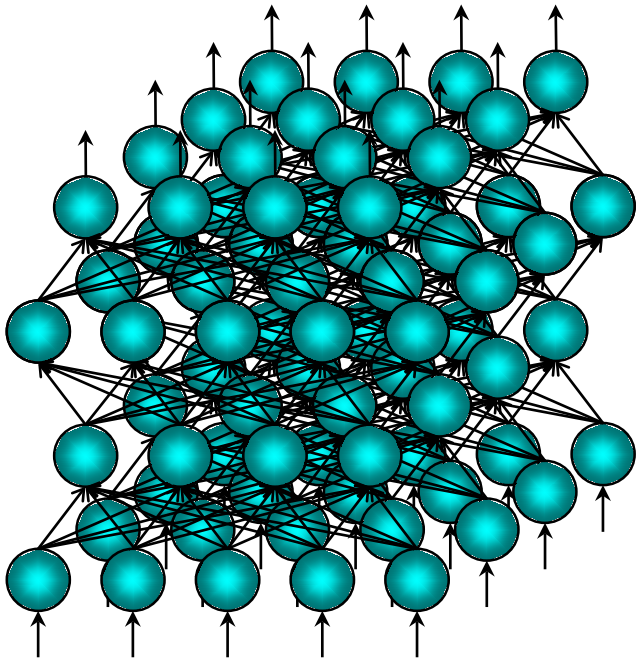
学習する
＝神経回路が
変化する

「学習」の構造

正しい答　y

誤差

神経回路の答　f(x,w)



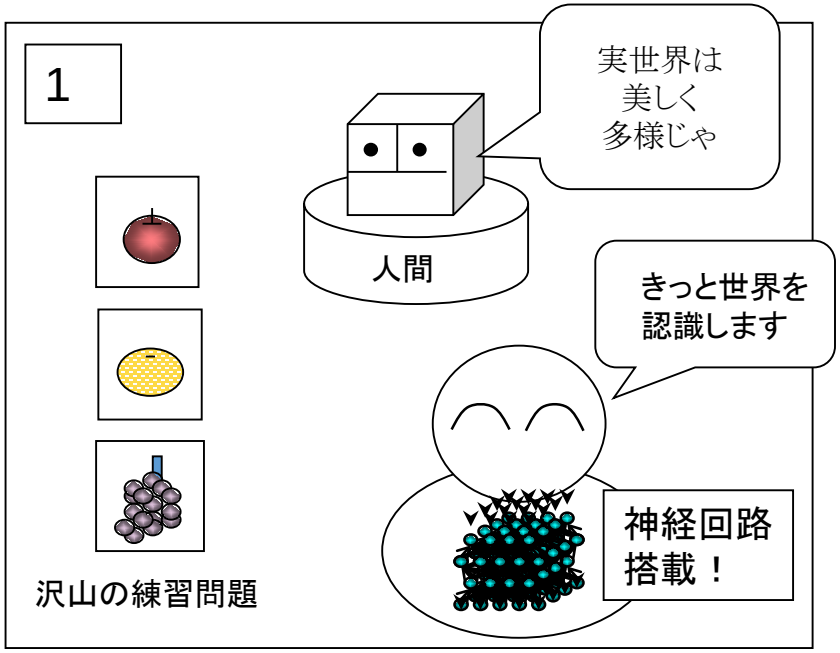
結合
荷重 w

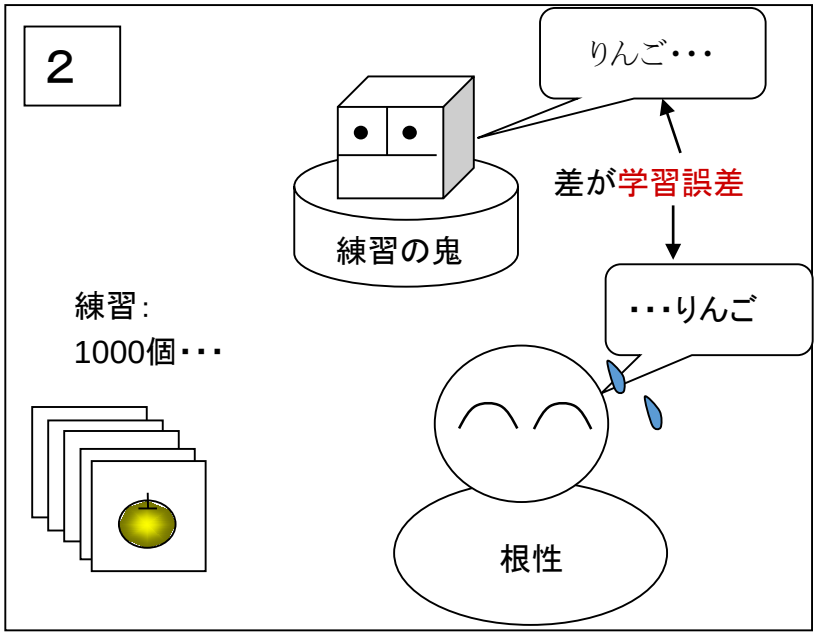
画像 x

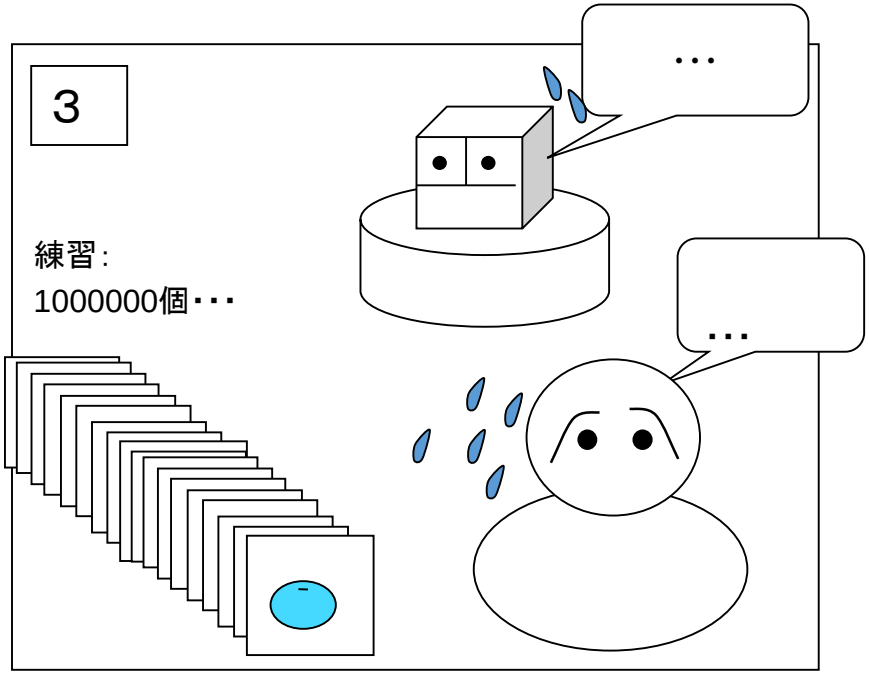
神経回路網の答えと
正しい答えの誤差は

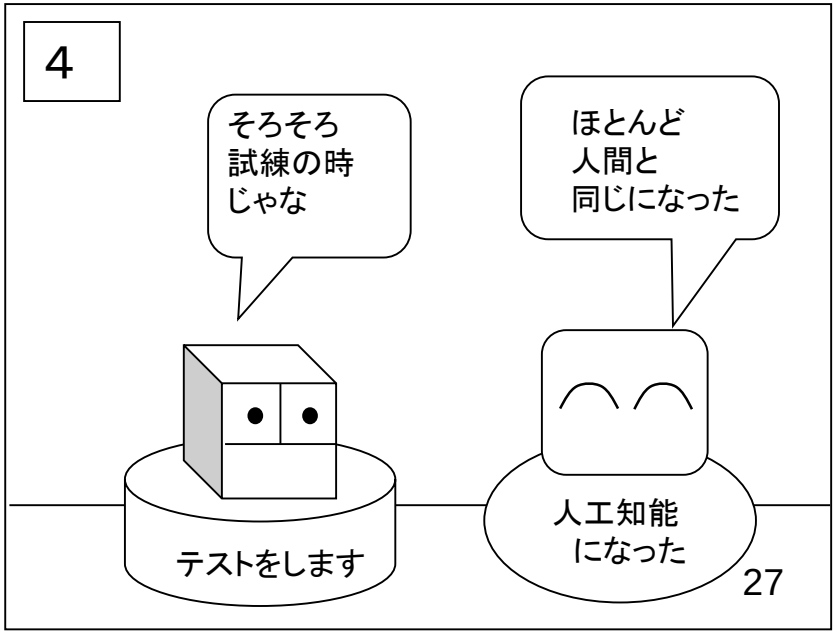
$(y-f(x,w))^2$

誤差が小さくなるように
w を変えていく

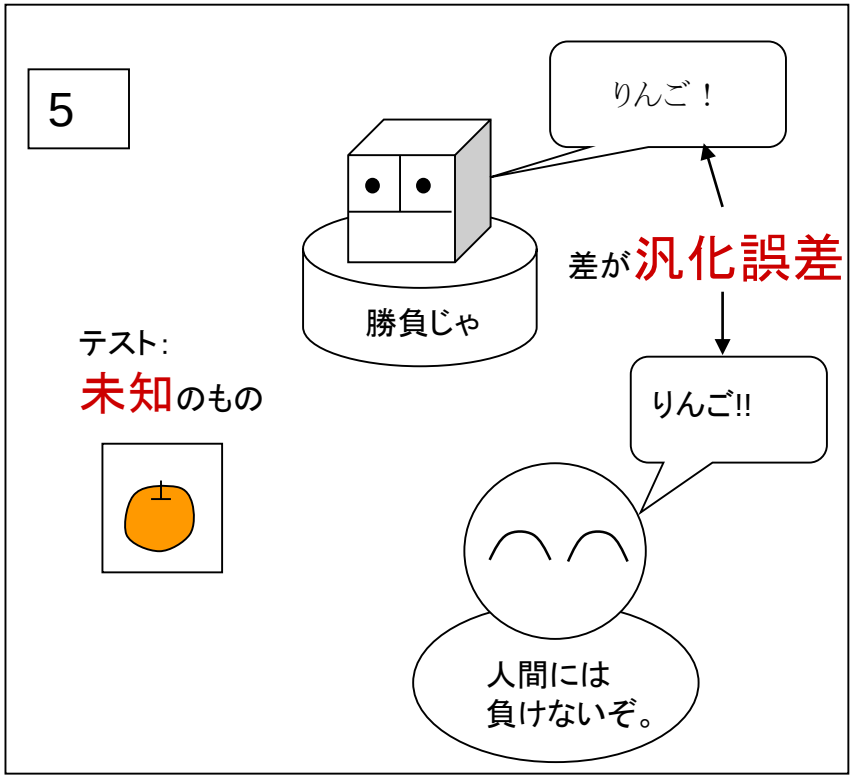

1
実世界は
美しく
多様じゃ
人間
きっと世界を
認識します
神経回路
搭載！
沢山の練習問題


2
りんご・・・
練習の鬼
差が学習誤差
・・・りんご
練習：
1000個・・・
根性


3
・・・
練習：
1000000個・・・
・・・


4
そろそろ
試練の時
じゃな
ほとんど
人間と
同じになった
テストをします
人工知能
になった

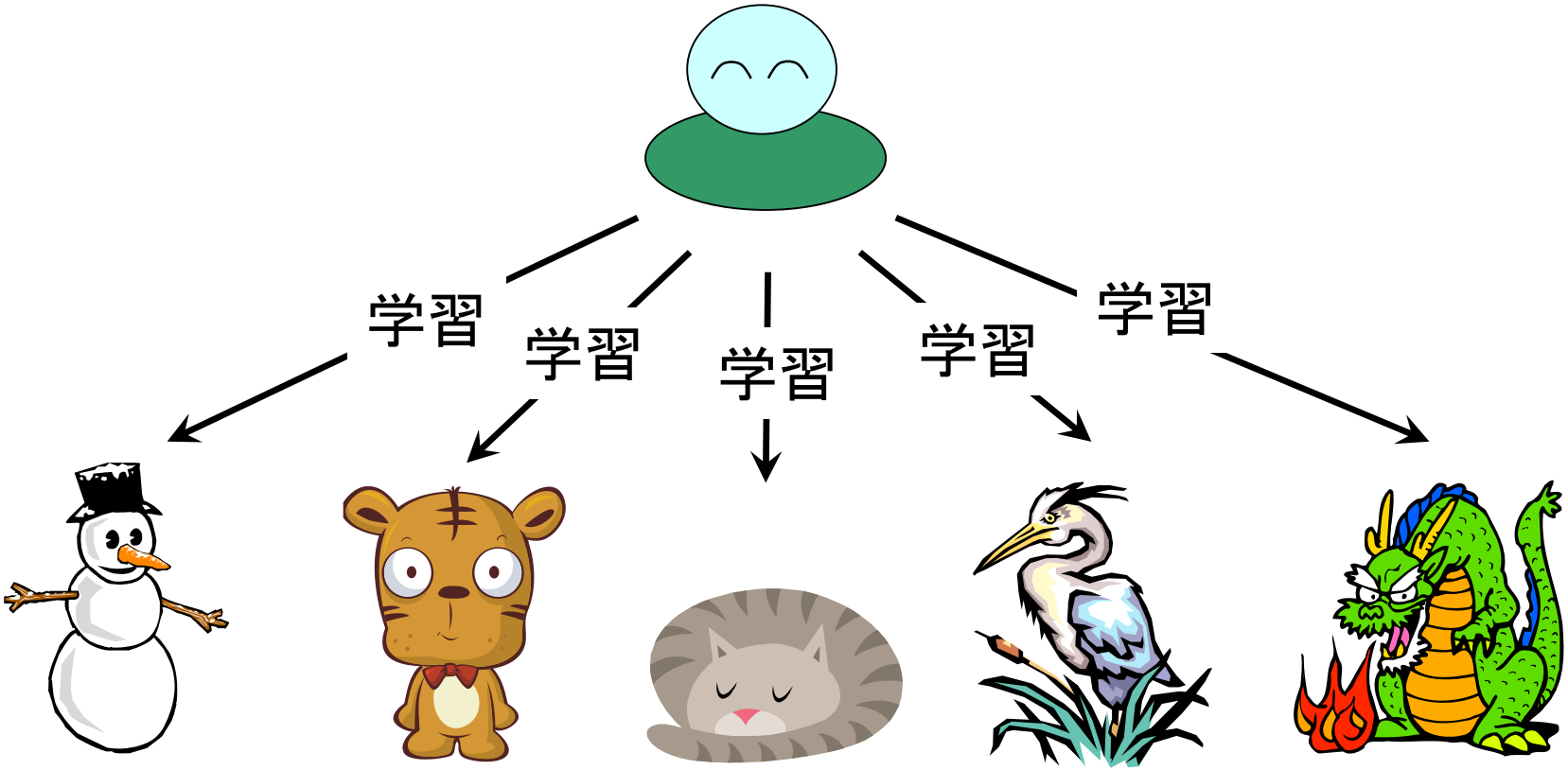
知らないものでも予測できる



◎　たくさんの例を学習すると神経回路網は
知らないものについても正しく答えられるようになってくる。

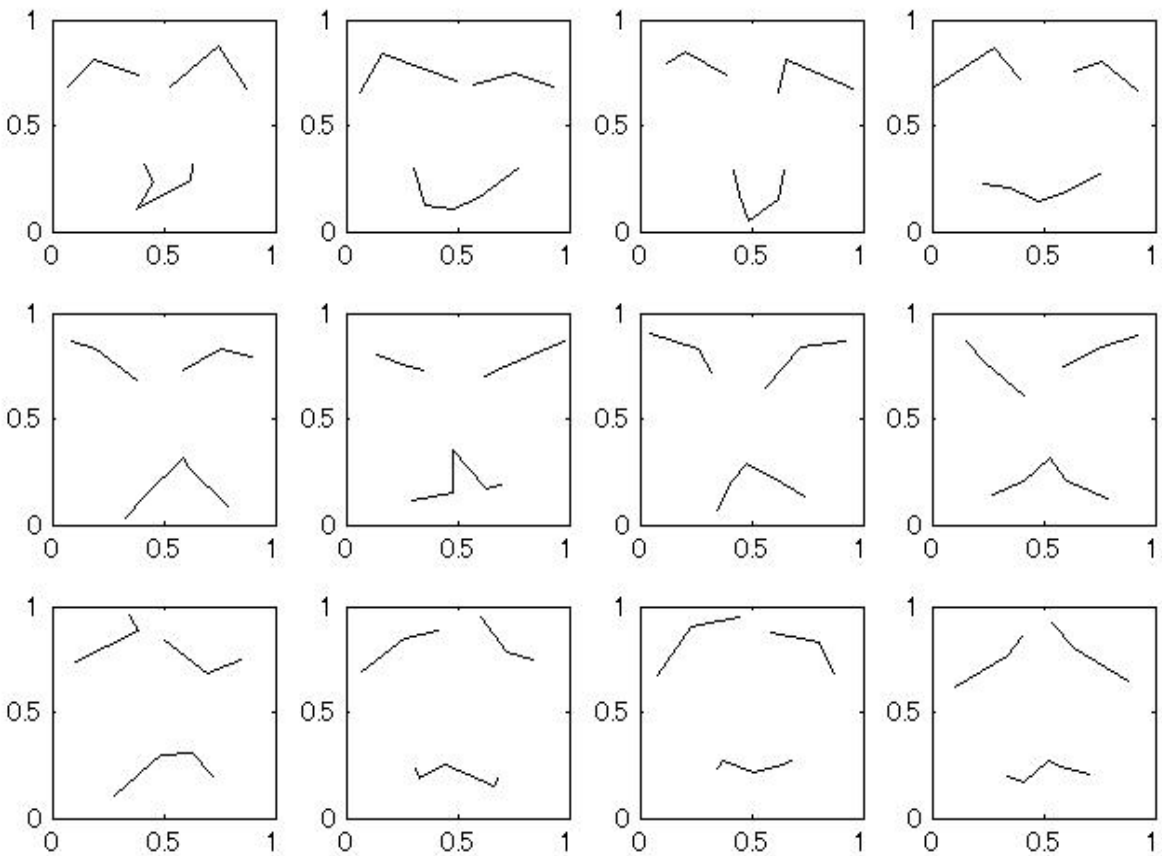
機械学習は何にでもなれる。

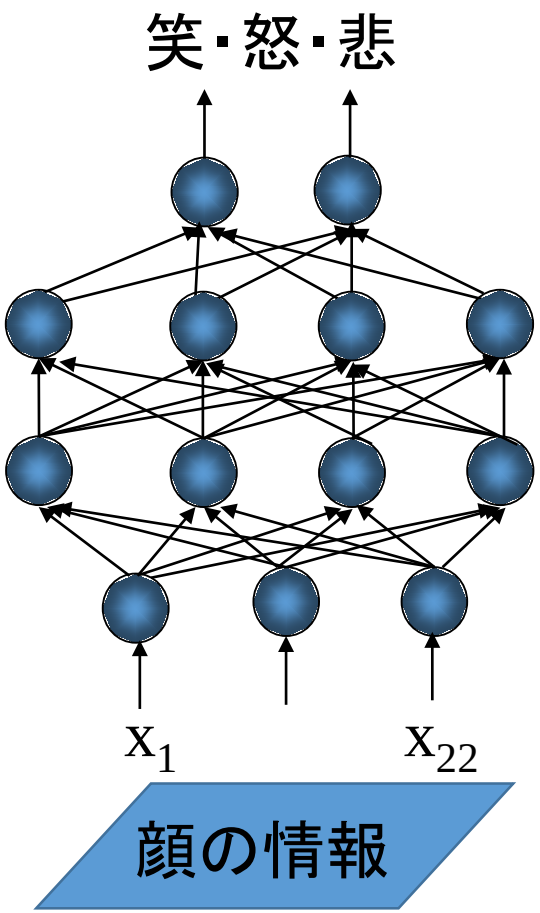
機械学習のモデルは、学習によってどんな関数になることもできる。しかしその数学的な性質は解明されていない。
まるで「召喚魔法」といわれるときもある。



顔から笑・怒・悲を読み取る

「顔から笑・怒・悲を読み取る人工知能」を作ってみよう。





学習してみた

問3

従来のデータ解析と機械学習とで異なる点は何か

新しい世界へ

実世界

観測データ

深層学習

代数幾何

学習理論

機械学習

微分幾何

経験過程

カーネルマシン

ベイズフィルタ

空間統計