# 情報学習理論

最後までご聴講いただきありがとうございました。

渡辺澄夫
東京工業大学

# 何処へ

世界中をめぐり学習モデルたちに出会った。ここから何処へ？

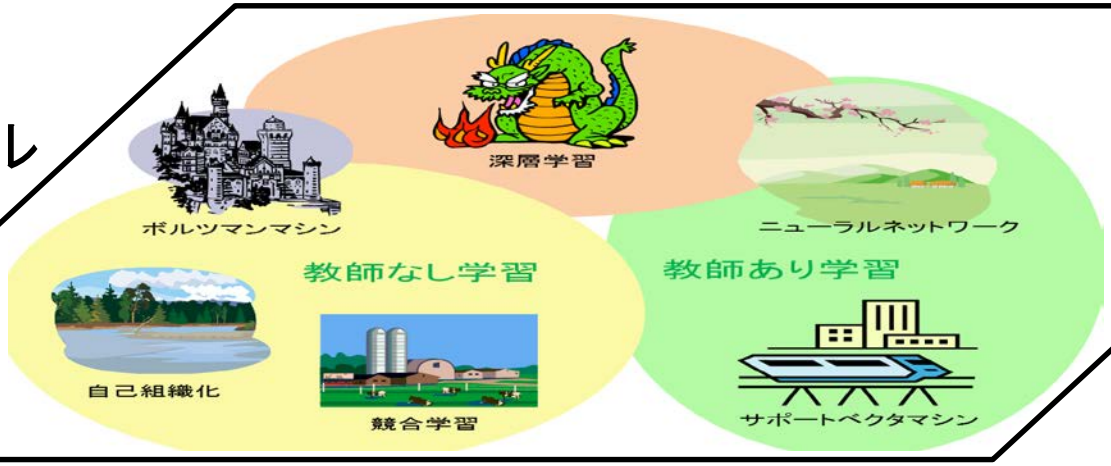
第3世界
実世界

物理学　心理学　情報　通信
化学　社会学　金融　交通
生物学　文学　医療福祉
環境学　経済学　エンターテイン

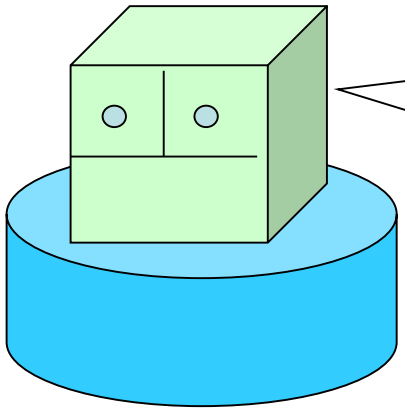
第2世界
学習理論

大数の法則　情報量規準
中心極限定理　交差確認法
経験過程　自由エネルギー
相対エントロピー　VC次元
汎化誤差　平均場近似

第1世界
学習モデル



# 学習理論とは

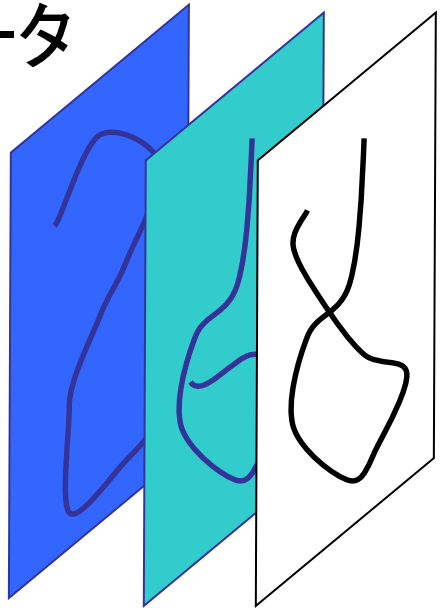


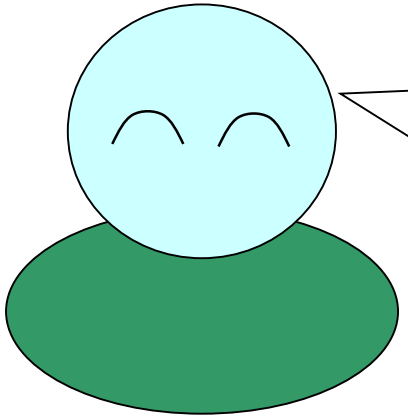
先生

n 個答えました。

データ

…





生徒

僕はどのくらい先生と似ているのだろう。

「先生と生徒の距離$^2$」 ＝ 確率変数 / n

# 学習理論とは

様々なデータがあり、様々な学習モデルがある。

真の分布はわからない。

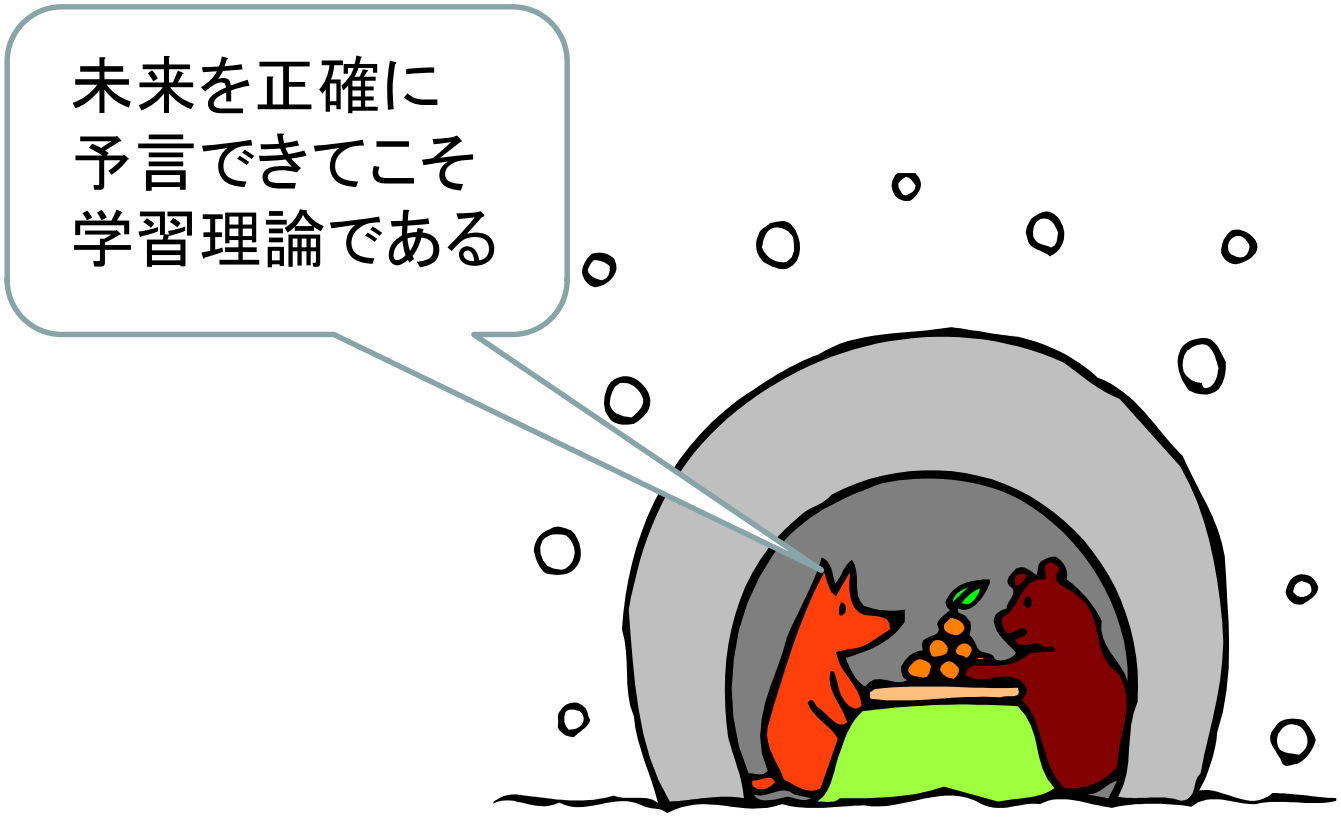
しかし「学習」という確率的現象においては

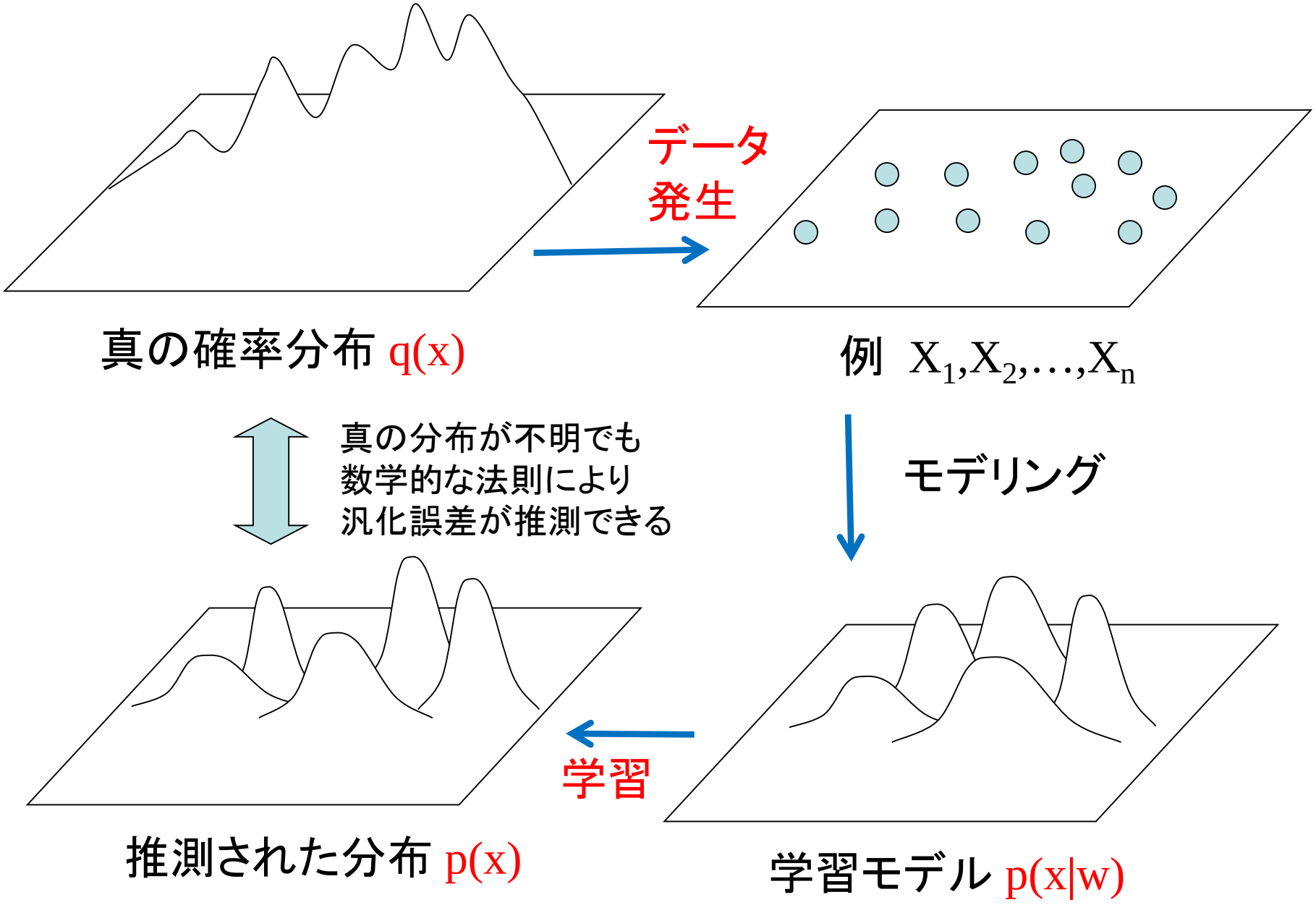
真の分布、データ、モデルによらない数学的な法則がある。

その法則に基づいて真の分布について推測を行うことができる。

それが学習理論である。

# 学習理論で何がわかるか（予測編）




データ
発生
真の確率分布 q(x)
例 $X_1,X_2,\ldots,X_n$
真の分布が不明でも
数学的な法則により
汎化誤差が推測できる
モデリング
学習
推測された分布 p(x)
学習モデル p(x|w)

# 汎化誤差の理論

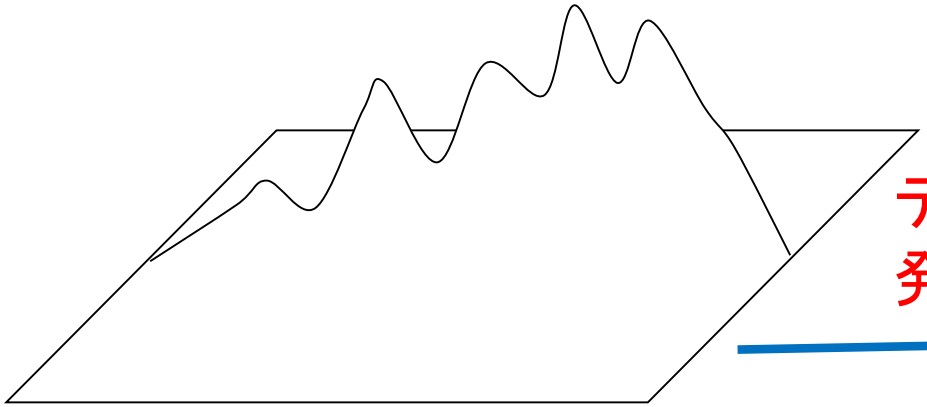
様々なデータがあり、様々な学習モデルがある。

真の分布はわからない。

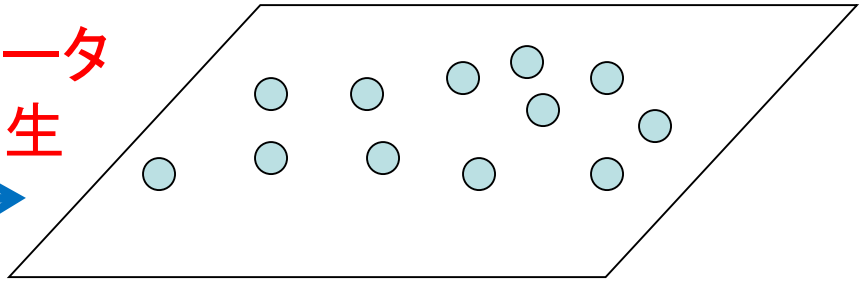
しかしながら、「汎化誤差 ー 学習誤差」について
真の分布、データ、モデルによらない数学的な法則がある。

その法則を応用して汎化誤差を推測することができる。

（例）情報量規準、クロスバリデーションなど

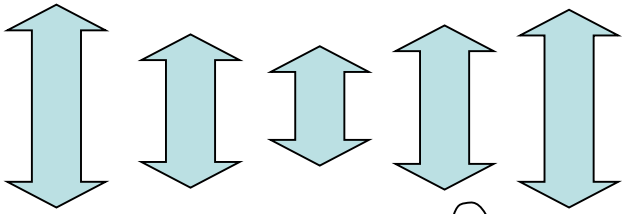


データ
発生



真の確率分布 $q(x)$

例 $X_1, X_2, \ldots, X_n$

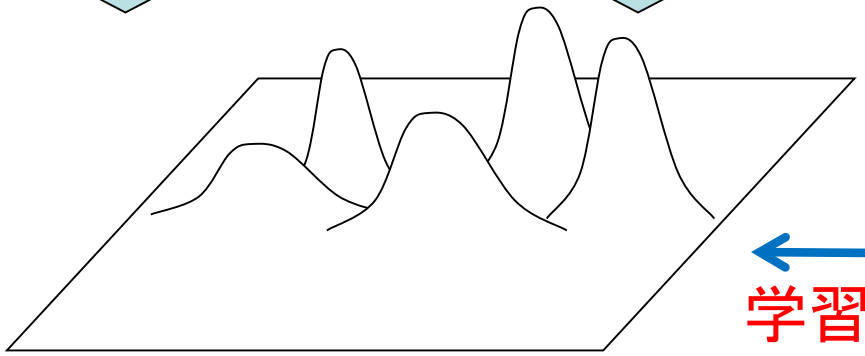


汎化誤差の
小さいモデル
を選びましょう

モデリング

候補のモデル

学習



推測された分布 $p_1(x), \ldots, p_5(x)$

# 汎化誤差によるモデルの評価

様々なデータがあり、様々な学習モデルがある。

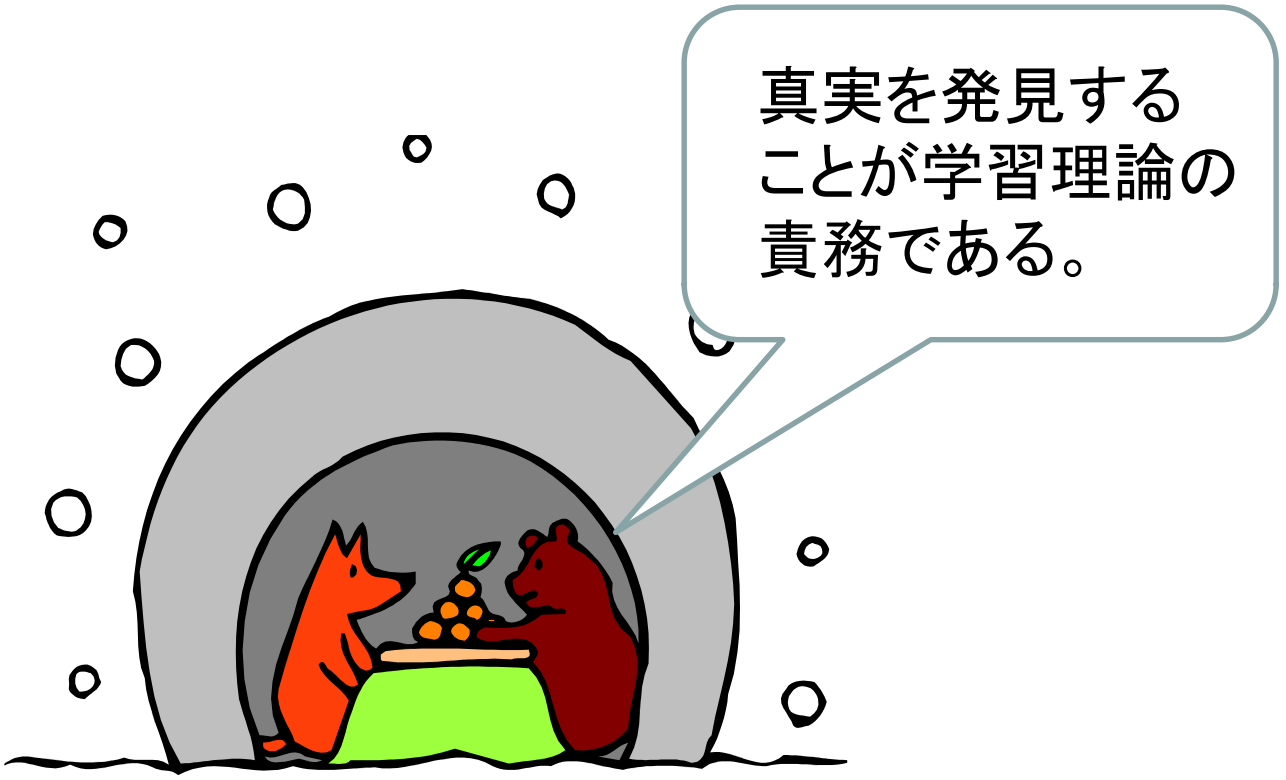
真の分布はわからない。

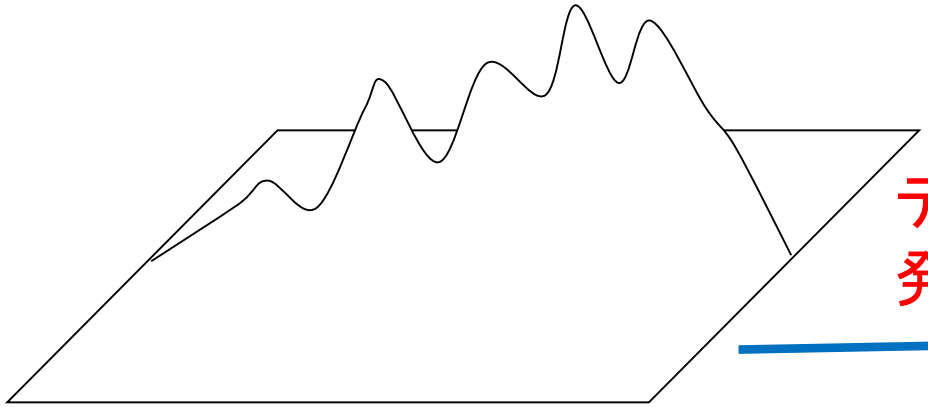
候補となる学習モデルが複数あるとき

汎化誤差が小さくなると期待されるものを選ぶことができる。

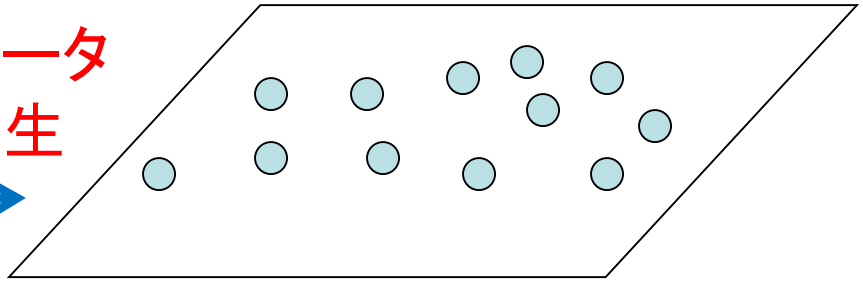
（注意）汎化誤差の推測には確率的揺らぎが含まれているので
必ず汎化誤差を最小にするモデルが見つかるわけではありませんが
データが与えられたとき、期待値として汎化誤差を最小にするモデルを
選ぶことができます。

# 学習理論で何がわかるか（確率編）





データ
発生



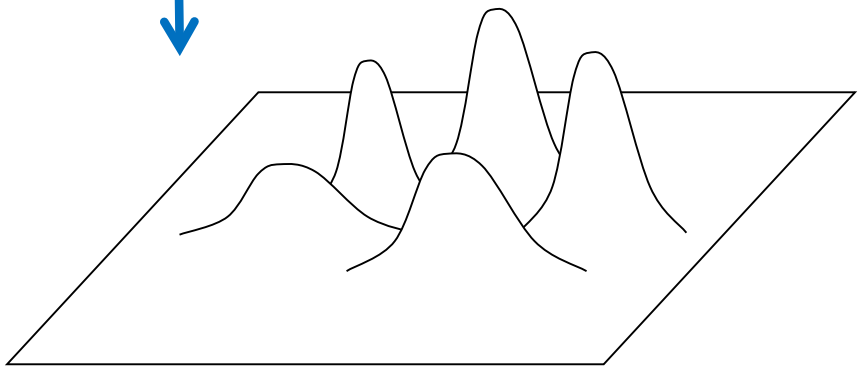
真の確率分布 $q(x)$

例 $X_1, X_2, \ldots, X_n$

モデリング

事後確率が計算できる

「データが与えられた
ときの学習モデルの
事後確率が大きい
ほど真の分布に近いだろう」



学習モデル $p(x|w)$

# 事後確率の理論

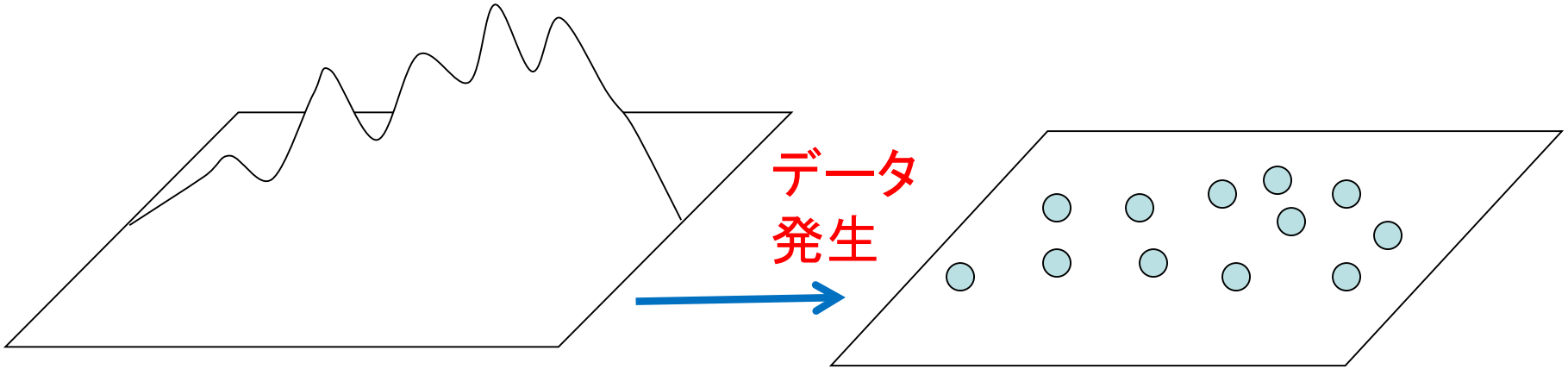
様々なデータがあり、様々な学習モデルがある。

真の分布はわからない。

しかしながら、モデルの事後確率について
真の分布、データ、モデルによらない数学的な法則がある。

その法則に基づいて事後確率を計算することができる。

（例）情報量規準、自由エネルギー（対数周辺尤度の符号反転）




真の確率分布 $q(x)$

例 $X_1, X_2, \ldots, X_n$

モデリング

候補のモデル

事後確率の大きいモデルを選ぶ

# 事後確率によるモデルの評価

様々なデータがあり、様々な学習モデルがある。

真の分布はわからない。

候補となる学習モデルが複数あるとき

事後確率が大きくなるものを選ぶことができる。

（注意）事後確率はあくまでも確率なので事後確率を
最大にするモデルが真のモデルと必ず一致する
わけではありませんが、データに対して学習モデルの
主観によらない評価を行うことができます。

# 学習理論を目指すみなさんに

「理論なのに数理が出てこないんですが・・・。」

「予測精度の解明」および「構造の発見」は学習理論において
中心的に重要な課題であり、上記のことについては

数学的な定義
定理の記述
定理の証明

を行うことができます。将来、必要になったら勉強してみてください。

# 免許皆伝

このページの内容は理解できなくても単位の履修上の問題はありません。
しかしながら、大学を卒業した後、実世界を生きていく上で大切なことを述べます。

(1)　現実的な状況では真の分布は無限に複雑であり、人間が用意する有限個のモデルの集合の中に真の分布とぴったりと一致するものはないと考えられます。データの数が多くなるにつれて少しずつ真の分布の詳しい理解ができるようになります。汎化誤差の推測値を最小にする方法はデータの持つ情報を最大限に予測に活かすために有効であると考えられます（赤池弘次）。そのことを数学的に証明できるモデル族があります（柴田里程）。

(2) 真の分布は不明ですが、仮に人間が用意したモデル族の中に真の分布とぴったり一致するものが含まれているという特別なケースを考えましょう。真の分布とぴったり一致するモデルの中で最もパラメータ次元の小さいものを【真のモデル】と呼ぶことにします。データの数が無限大に近づく極限を考えましょう。汎化誤差の推測値を最小にするモデルを選んでも、【真のモデル】が選ばれる確率は1には近づかないことが知られています。一方、事後確率を最大化する方法(I.J.Good)でモデルを選ぶと、【真のモデル】が選ばれる確率が1に近づきます。

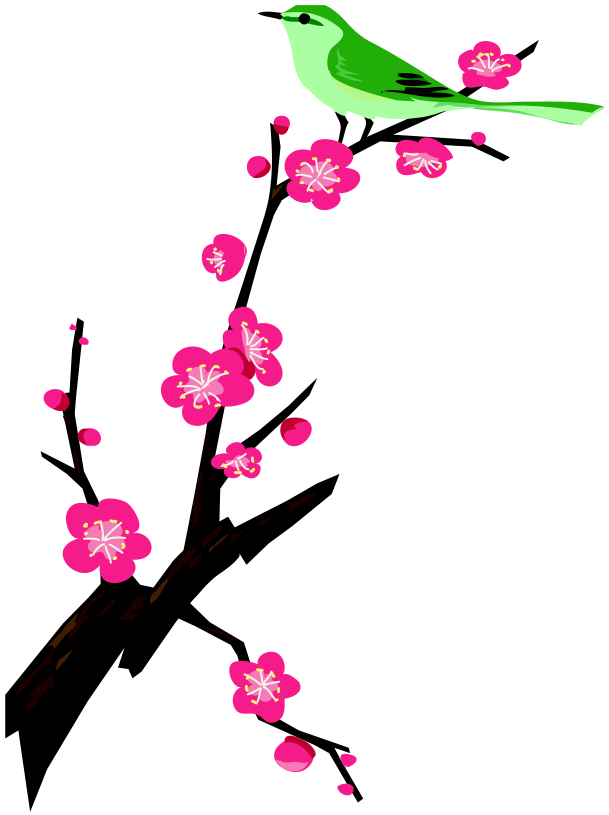
以上の(1)(2)はどちらも正しい記述ですので、どちらが正しいかという論争は無意味です。
また哲学あるいは心理学の対象となる課題でもありません。(1)も(2)も数学的な事実を述べたものであり、真の分布がわかることのない現実の世界を生きていく私たちが行うモデリングの強力な基礎になります。

# 問1

次の表を作りましょう。

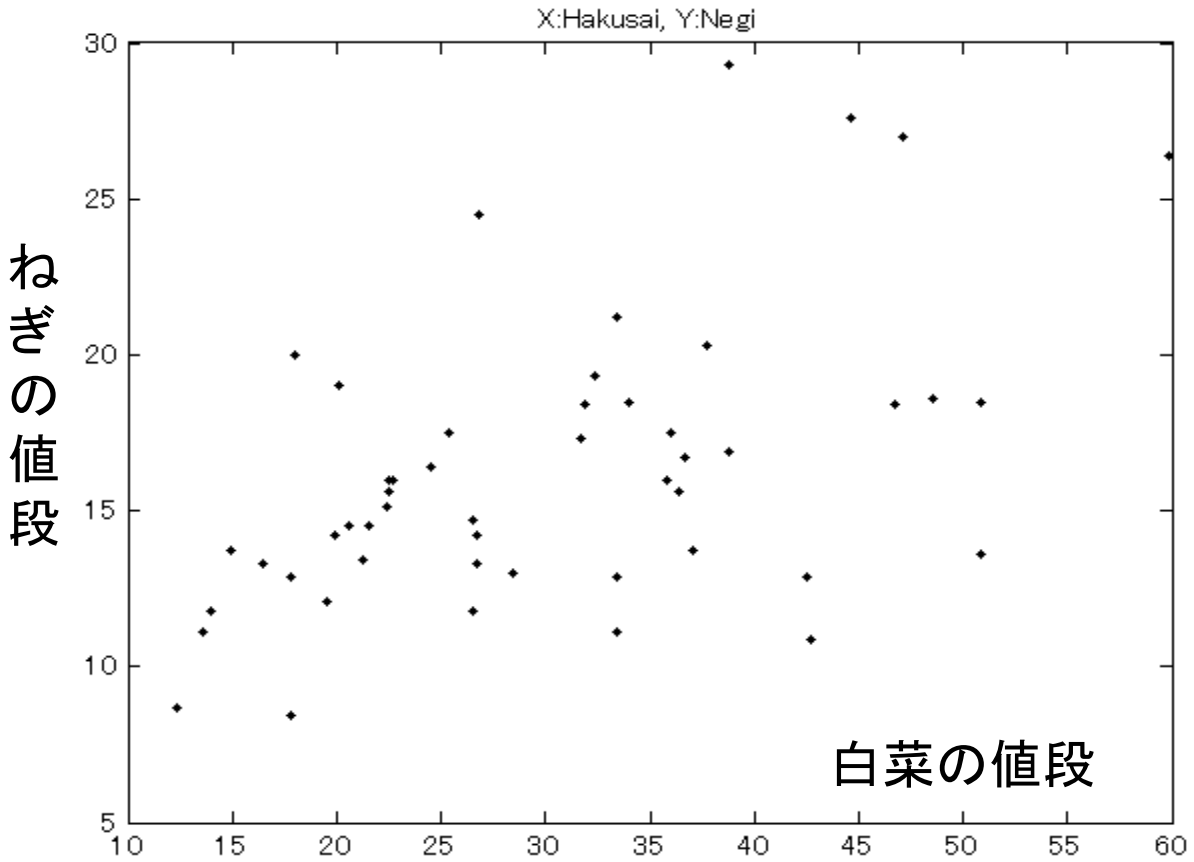
| | 何を測るか | 例 |
|---|---|---|
| できるだけ正確に<br>予測したい | | |
| できるだけ確率を<br>大きくしたい | | |

# 春が来る



# 実データに向かって

時系列予測の例：1970年1月から2013年12月までの白菜とねぎの値段
「政府統計の総合窓口」のデータを使用しています。
http://www.e-stat.go.jp/SG1/estat/eStatTopPortal.do




白菜の値段からねぎの値段を知りたいとき、
どんなモデルを使うとよいのだろうか。(n=50)

# いろいろなモデル

真の関係はわからない。多項式モデルを複数考えてみる。

$Y=a+$雑音

$Y=a+bX+$雑音

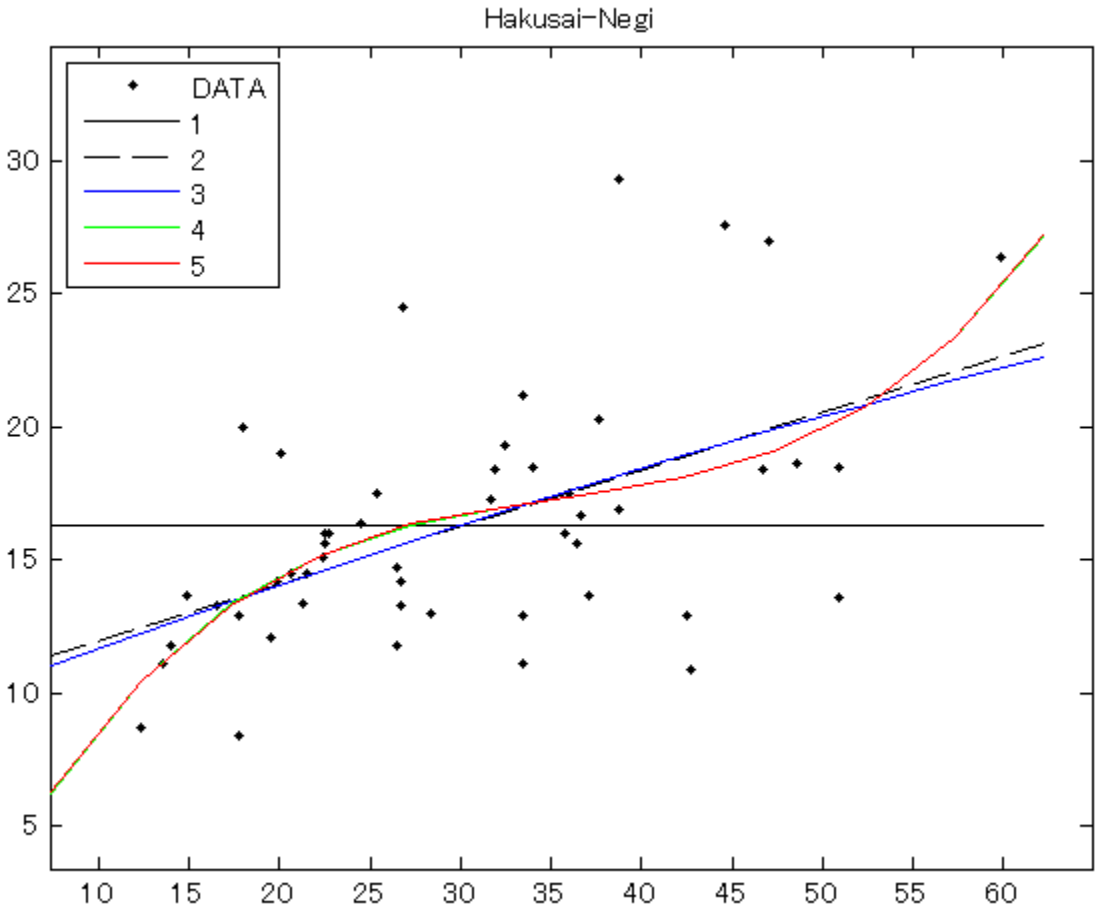
$Y=a+bX+cX^2+$雑音

$Y=a+bX+cX^2+dX^3+$雑音

$Y=a+bX+cX^2+dX^3+eX^4+$雑音

二乗誤差を最小にするパラメータは簡単に求められるが・・・。

# 学習してみた




グラフを見てもどれが良いのかわからない・・・。
高次元だとグラフを見ることもできない・・・。

# 情報量規準

様々なデータがあり、様々な学習モデルがある。

AIC(赤池情報量規準）　：　汎化誤差の推測値
BIC(ベイズ情報量規準）：　自由エネルギーの近似値

DIM[1]: AIC=1095.792275, BIC=1136.084199
DIM[2]: AIC=827.373038, BIC=885.964012
DIM[3]: AIC=857.439524, BIC=945.266886
DIM[4]: AIC=856.379831, BIC=969.305206
DIM[5]: AIC=885.903381, BIC=1027.059018

モデルを客観的に比較することができる。

(注) 選ばれるモデルは、データの数にも依存します。

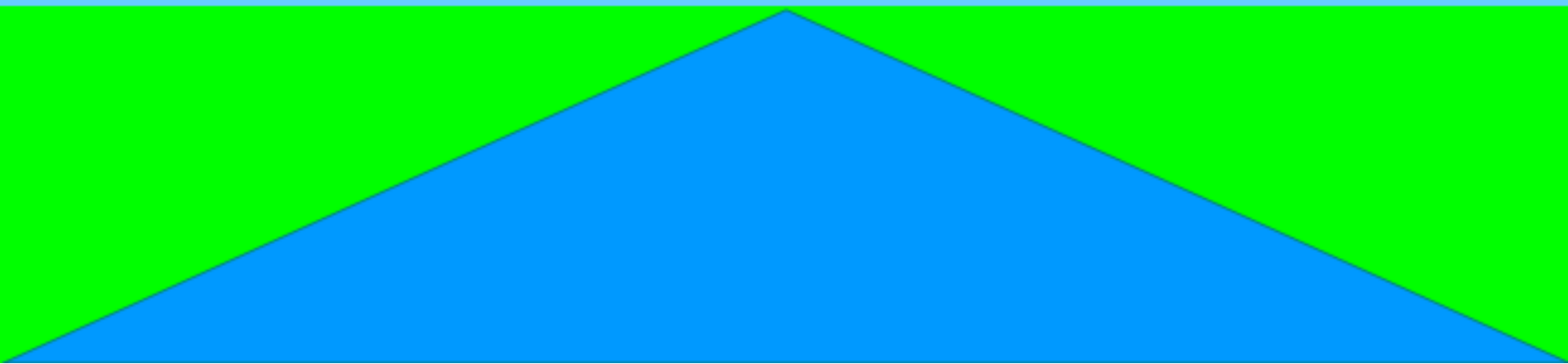
(注) AIC, BICは、構造持つモデルの評価には使えません。例えば、神経回路網、深層学習、SVM、混合正規分布、K-means法、自己組織化写像、ボルツマンマシン、隠れマルコフモデルにはAIC,BICは使えません。それらのモデルの汎化誤差や事後確率を求める理論については現在研究が行われています。

# 問題2

次の表を作りましょう。

| | n=20 | n=25 | n=30 | n=60 | n=120 |
|---|---|---|---|---|---|
| AICを最小にするモデル | | | | | |
| BICを最小にするモデル | | | | | |

水源は不明でも、それでも川は流れている。（ポアンカレ）



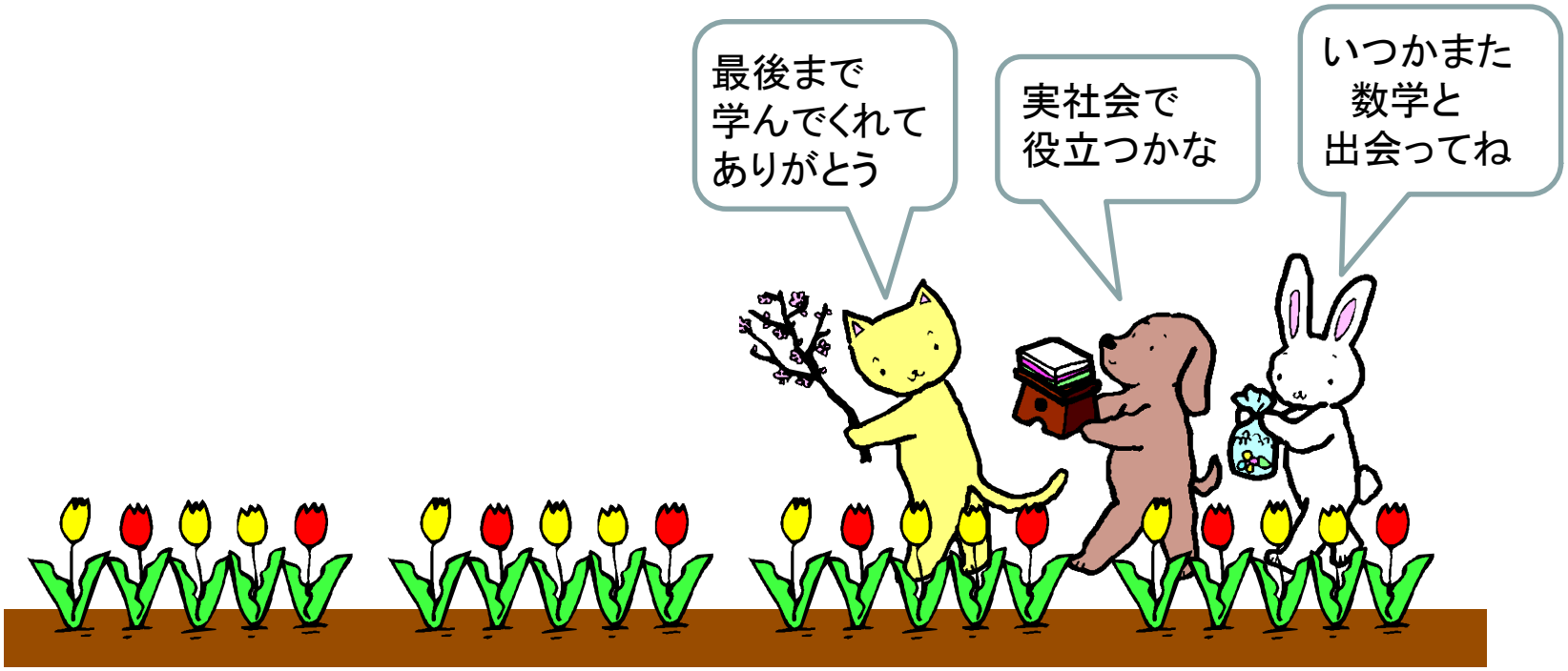
# 学習理論と社会

学習理論は、もともと人工知能を作るため、すなわち音声・画像の認識や自然言語・時系列の理解に用いることを目的として研究されてきました。

今日では、膨大データ、巨大ネットワーク、深い学習モデルが実現され、社会経済、自然科学、芸術文化のありかたを変革し始めています。
私たちが気づかないところで既に学習モデルが活躍しています。

学習理論は科学あるいは技術なので、みんなの幸せに役立つかどうかは、使い方しだいです。

今度こそ人類は科学あるいは技術を賢く正しく使って欲しいと思います。


最後まで
学んでくれて
ありがとう
実社会で
役立つかな
いつかまた
数学と
出会ってね