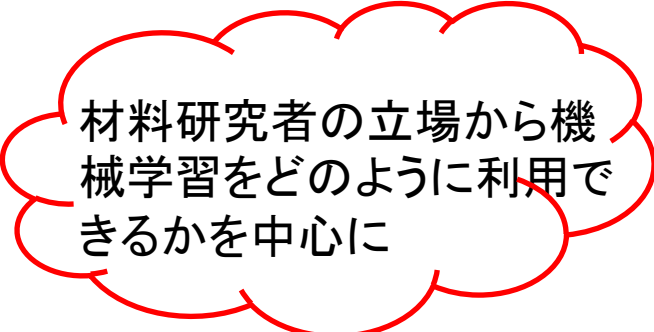


材料研究者のための機械学習入門

滋賀大学 徳田陽明
(tokuda@edu.shiga-u.ac.jp)

本日の流れ

- **機械学習の基礎**
 - 機械学習とは？
 - 手法の概要
- **少し専門的な数学の説明**
 - 線形代数(エクセル使用)
 - 統計学
- **実装方法の概略(エクセル使用)**
- **実際の材料への適用例**
 - ガラス材料に応用する場合の課題
 - 世界の研究動向
 - 光学ガラスへの適用の実際例



材料研究者の立場から機械学習をどのように利用できるかを中心に

機械学習とは？

Wikipediaによると・・・

機械学習（きかいがくしゅう, (英: machine learning)とは, 人工知能における研究課題の一つで, **人間が自然に行っている学習能力と同様の機能をコンピュータで実現**しようとする技術・手法のことである。

<https://ja.wikipedia.org/wiki/機械学習>

AIが囲碁のトップ棋士に勝った！



連載講座「IGOサイエンス」特別座談会 & 全5局解説
新時代到来！ ～アルファ碁の驚異の序盤力と弱点に迫る～
 王銘琬九段×マイケル・レドモンド九段×大橋拓文六段

●第40期 棋聖戦七番勝負 第4局 井山裕太 vs 山下敬吾
井山 疾風4連勝 自戦解説 井山裕太棋聖
 2大好評連載 漫画 いやあせつらストーリー **井山裕太物語**
 天棋TENKI 第2部 松田一輝

●第17回 農心辛ラーメン杯世界最強戦・第3ラウンド
 ●第54期 十段戦五番勝負 第1・2局 ●第28期 女流名人戦三番勝負 第1・2局

好評連載
 「私と囲碁」 2次ノブリテック 水野雅生
 囲碁大使 戸島花の ドリーム囲碁サロン 戸島花
 教えてアツシ先生! 人気定石Q&A 講師・加藤充志九段
 シンプル is ベスト 石の強弱でわかる次の一手 講師・熊幸六段
 エッセイ&パズル ハッピーワールド

日本棋院90年史 「小林光一、ライバルを破って頂点に」
 碁会所70年 坂田栄男・藤沢秀行も 碁会所で育った
 大矢布石研究所 所長・大矢浩一九段
 創作詰碁にチャレンジ 出題・青木紳一九段
 トライアル50
 「井山裕太物語」 大和正樹
 「天棋—TENKI—」 松田一輝

別冊付録 次の一手
 「7路で上達! ⑧～ヨセ編～」 出題・水間俊文七段
 あなたの「碁ライフ」が充実する

碁ワールド 4月20日発売 定価885円(税込) **5月号 2016 May**
 Google 〇お求めはお近くの書店でご注文ください 〇日本棋院

年間購読(各種会員特典有り)をご希望の方は日本棋院普及事業部まで 〒102-0076 千代田区五番町7-2 TEL. 03-3288-8723

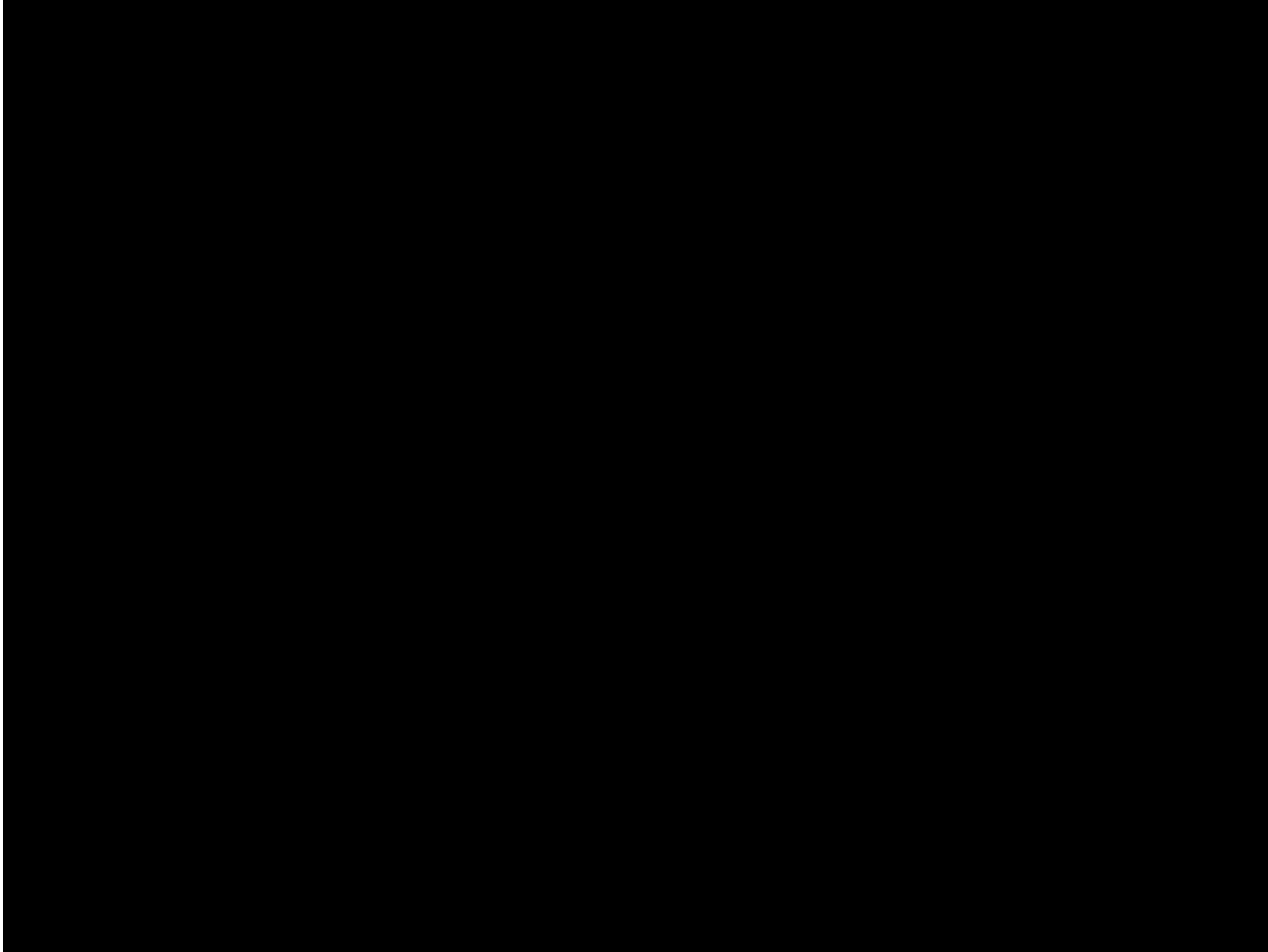
世界に衝撃! 最強AI
 グーグルデイトープマインド
 チャレンジマッチ

アルファ碁 VS 李世乭九段
 (人工知能) イ・セドル

https://www.nihonkiin.or.jp/publishing/go_world/goworld_201605.html

アルファ碁ゼロは、**AI同士の対戦で学習**し、プロ棋士に勝利

ブロック崩し



<https://www.youtube.com/watch?v=eG1Ed8PTJ18>

機械が自力で顔を認識した！

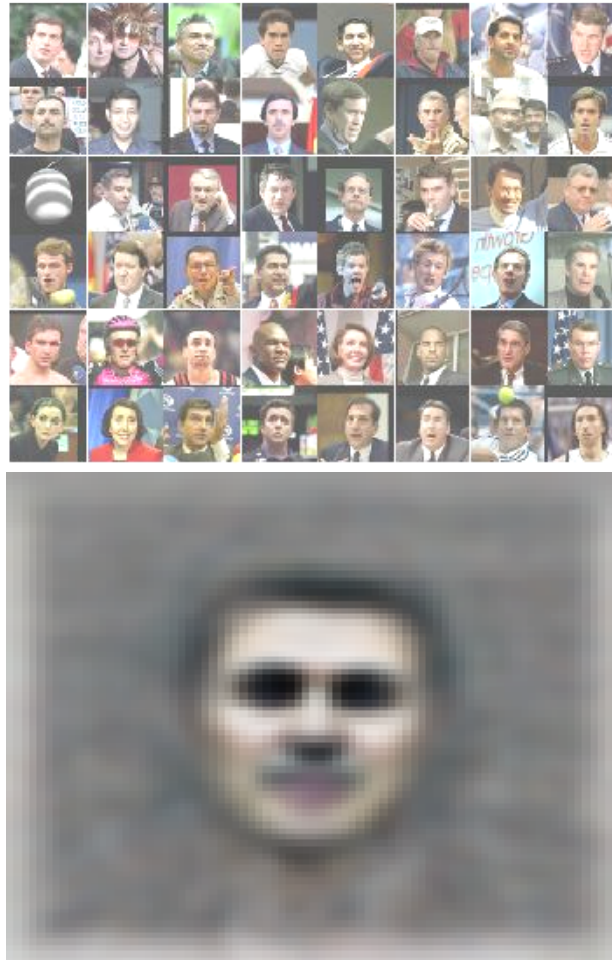


Figure 3. Top: Top 48 stimuli of the best neuron from the test set. Bottom: The optimal stimulus according to numerical constraint optimization.

レンブラント？



<https://www.nextrembrandt.com>

なぜ機械学習が注目されているのか

- GAFAによる実装
- データ量の増加
 - インターネット利用(画像など)の拡大
 - 実験データの蓄積
- コンピュータの性能の向上
 - ゲーム機が昔のスパコン並みの能力
- 解析アルゴリズムの充実
 - 統計計算, 行列計算のパッケージ化

機械学習は魔法の技術？

自動化, 自動認識などのニーズに見合うだけの性能を達成可能

機械学習とは？

機械学習とは、ある目的のためにデータを基にしてコンピュータ（機械）を用いて関係性を発見（学習）すること。そして、これを用いて、未知のデータを予測すること。

アルファ碁：強い碁アルゴリズムを作るために、過去の対局から着手の評価関数を作成した。そして、ある局面での最善手を見出した。

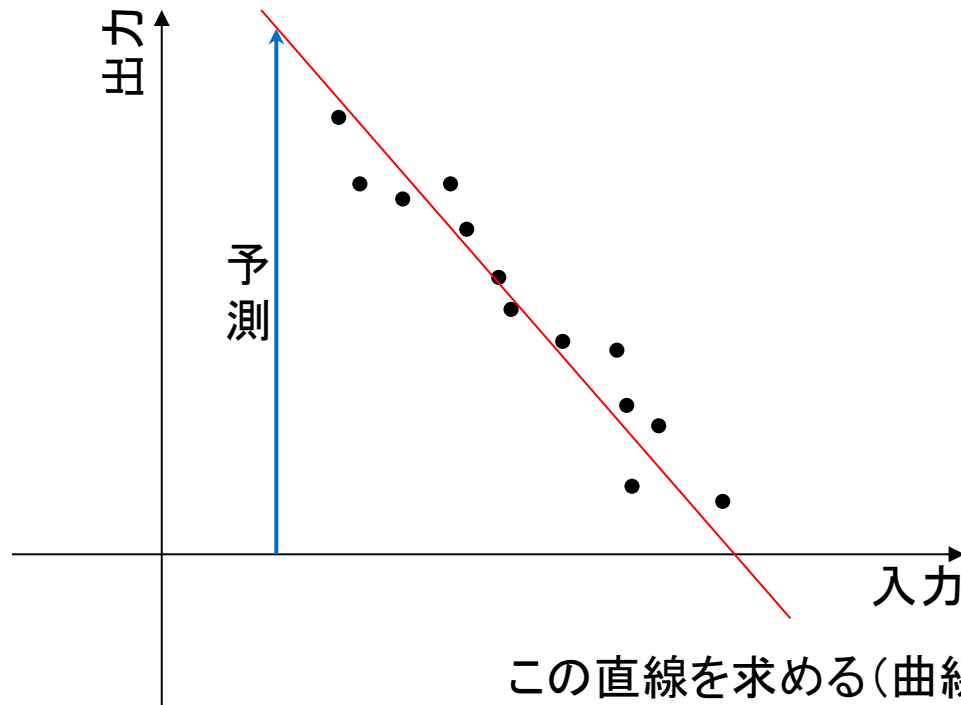
顔認識：顔を認識するために、ネット上の画像データを処理し、顔に特有の特徴を発見した。そして、ある写真が顔かどうかの判定を行なった。

機械学習でできること

- 回帰
 - 最小二乗法によるフィッティング
 - 重回帰分析
 - カーネル多変量解析
- 判別
 - 数字の判別
- クラスタリング
 - グループ分け
- パターンの発見
 - レコメンド機能 (Amazonなど)

回帰

回帰の例：消費者の属性と商品の販売数



回帰のためには・・・

データ点(●)と直線(—)との距離が最小になるようにする

入力

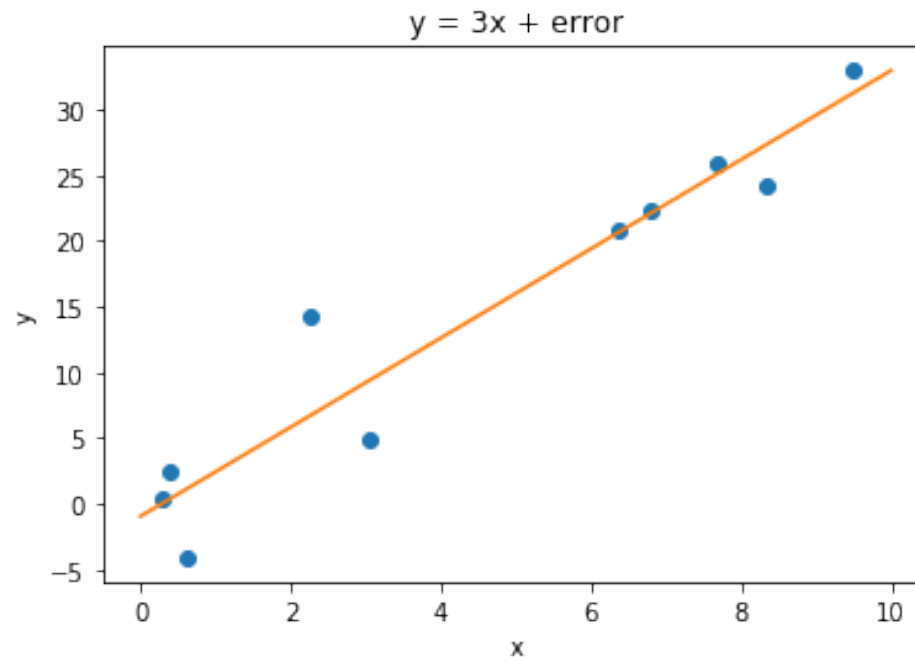
一次元：最小二乗法(単回帰)

多次元：重回帰

回帰

- 最も簡単な例

測定値10点を $y = 3x$ で線形回帰(フィッティング)

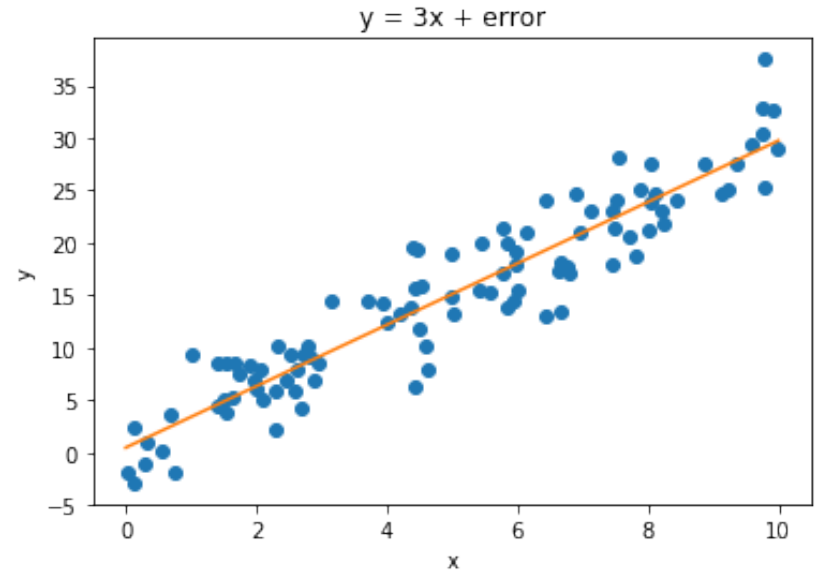


標準偏差: 0.3639

データ点が増えると・・・(10→100)



標準偏差: 0.3639

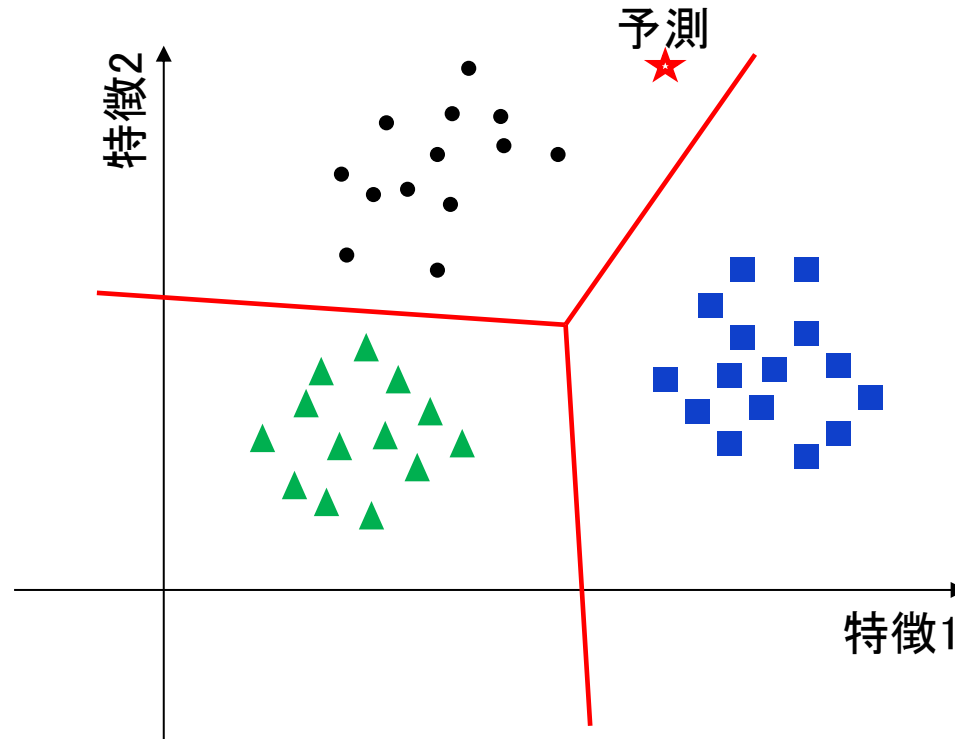


標準偏差: 0.1084

データ数の増加 → 予測の精度の向上

判別

判別の例: 動物の種類を見分ける



この直線を求める(曲線でも可)

判別のためには・・・

データ点と直線(—)との距離が最大になるようにする

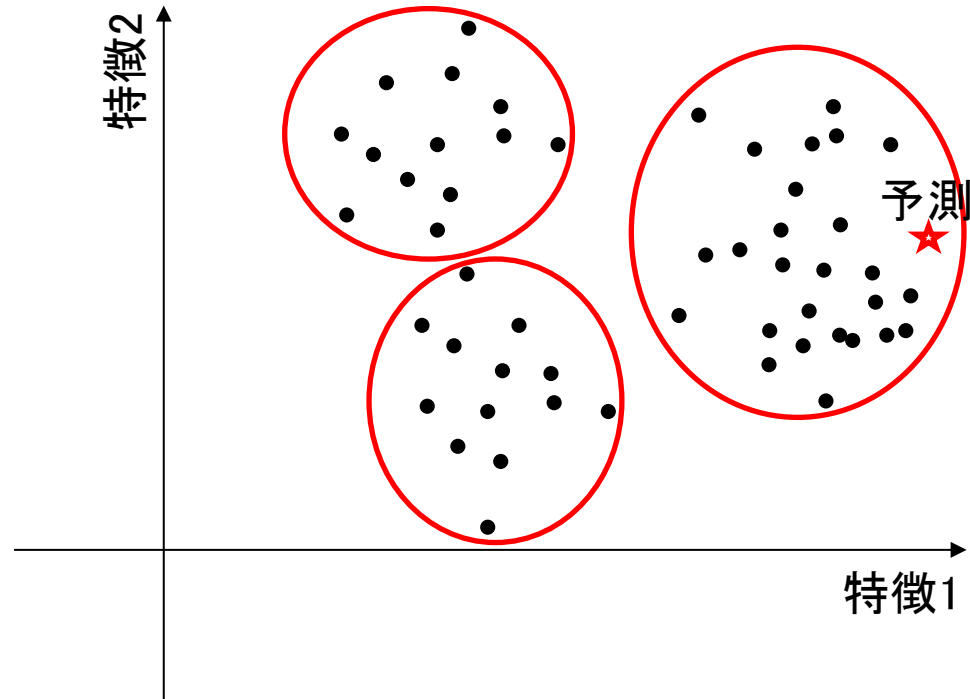
特徴

うまく設定する必要がある

高次元でも良い

クラスタリング

クラスタリングの例：居住地と消費行動の相関を見つける



クラスタリングのためには・・・

データ点と曲線(ー)との距離が最小になるようにする

(選んだ)特徴

データの素性が分かっていなくても良い

機械学習でできること

- 回帰

- 最小二乗法によるフィッティング
- 重回帰分析
- カーネル多変量解析

材料科学で役立つのは主にこれ？

- 分類

- 数字の判別

- クラスタリング

- グループ分け

- パターンの発見

何らかの関係が発見できた



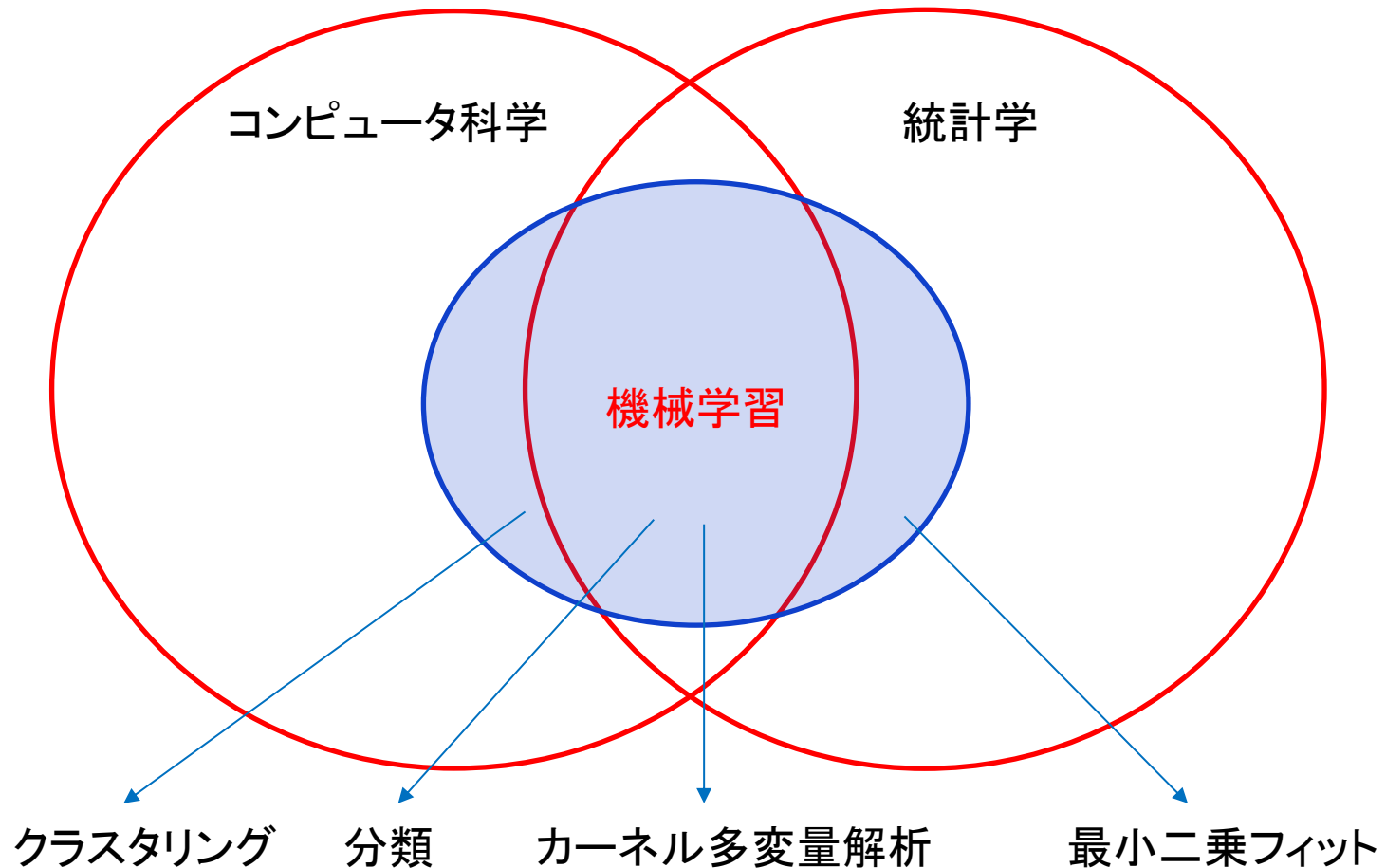
数式で表現できるはず

|評価値 - 予想値| → 最小

$$A_1 > A_2 > \dots > A_k$$

数学やアルゴリズムで実装可能！

機械学習の位置付け



機械学習は、統計学(数学)とコンピュータ科学の境界領域

本日の流れ

- **機械学習の基礎**
 - 機械学習とは？
 - **手法の概要**
- **少し専門的な数学の説明**
 - 線形代数(エクセル使用)
 - 統計学
- **実装方法の概略(エクセル使用)**
- **実際の材料への適用例**
 - ガラス材料に応用する場合の課題
 - 世界の研究動向
 - 光学ガラスへの適用の実例

なぜ機械学習？（特に回帰）

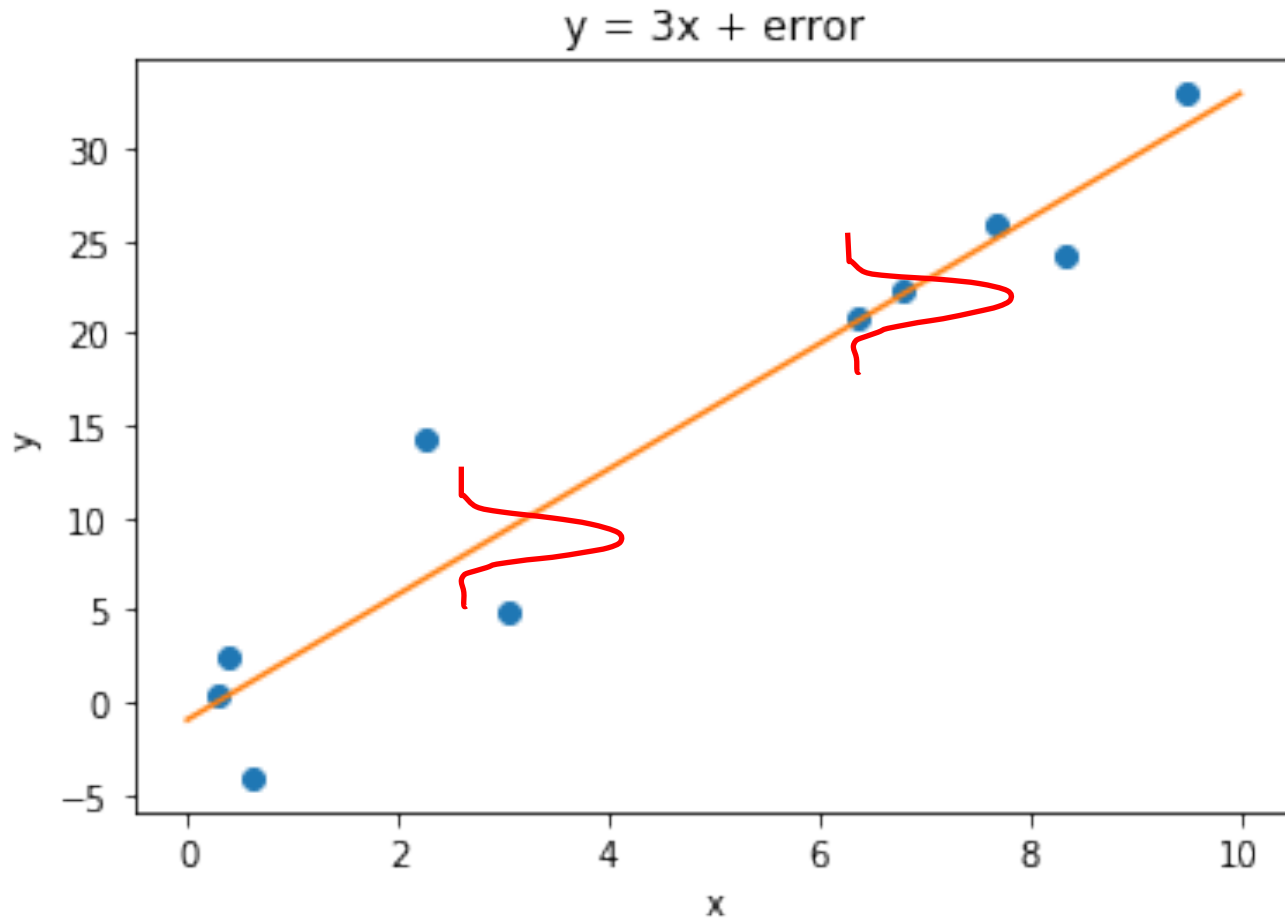
■ 回帰分析ならば、機械学習は不要？

- ✓ 最小二乗法で十分なのでは？
- ✓ 重回帰分析と何が違うのだろう？



- ◆ 統計的な(=誤差を含んだ)分析が可能
 - ✓ 最尤推定
- ◆ 非線形な(=多項式ではない)分析が可能
 - ✓ カーネル法
- ◆ 未知の情報に対する予測が可能
 - ✓ ベイズ推定

確率的な考え方の導入

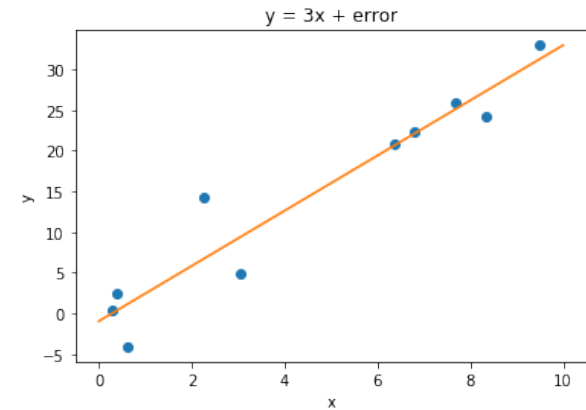


実測のデータは、図のy軸に沿った線上に
ガウス分布している(と考える)

与えられているデータ点(訓練データ)

$$(x_1, t_1), (x_2, t_2) \cdots (x_n, t_n)$$

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \rightarrow \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{pmatrix}$$



求めたい回帰曲線

$$y = w_0 + w_1 \cdot x = (w_0 \quad w_1) \begin{pmatrix} 1 \\ x \end{pmatrix} = \vec{w}^T \cdot \vec{x} = y(\vec{w}, \vec{x})$$

このように表記しておく利点

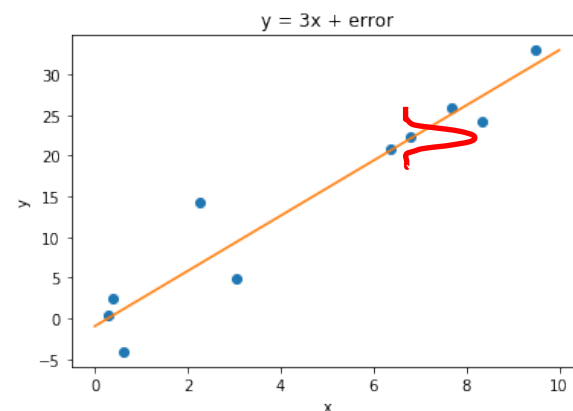
- ・多次元化(x^2, x^3 など), 多成分化(重回帰)が容易
- ・カーネル法への拡張が可能

ある値 x に対する t の値は、平均値 $y(\vec{w}, \vec{x})$ のガウス分布に従う

その確率は

$$p(t|x, w, \sigma^2) = \mathcal{N}(t|y(\vec{w}, \vec{x}), \sigma^2)$$

と表せる。



訓練データ(この例では10点の測定点)の発生確率は

$p(t_1|x_1, w, \sigma^2) = \mathcal{N}(t_1|y(\vec{w}, x_1), \sigma^2)$ などの積で表せる

$$p(\vec{t}|\vec{x}, \vec{w}, \sigma^2) = \prod_{k=1}^n \mathcal{N}(t_k|y(\vec{w}, x_k), \sigma^2) \quad \leftarrow \text{尤度関数}$$

訓練データの発生確率が求まった

最尤推定: 訓練データが得られる確率が**最も高く**なるように $\vec{w}$ を求める

尤度関数は $\exp(-x^2)$ の形をしているため、小さな値となりやすく、扱いに注意が必要

$$p(\vec{t}|\vec{x}, \vec{w}, \sigma^2) = \prod_1^n \mathcal{N}(t_k|y(\vec{w}, x_k), \sigma^2)$$

↓ 対数をとる

$$\ln p = -\frac{1}{2\sigma^2} \underbrace{\sum_1^n \{y(\vec{w}, x) - t_n\}^2}_{\text{これを最小にする(最小二乗法と一致)}} - \frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln(2\pi)$$

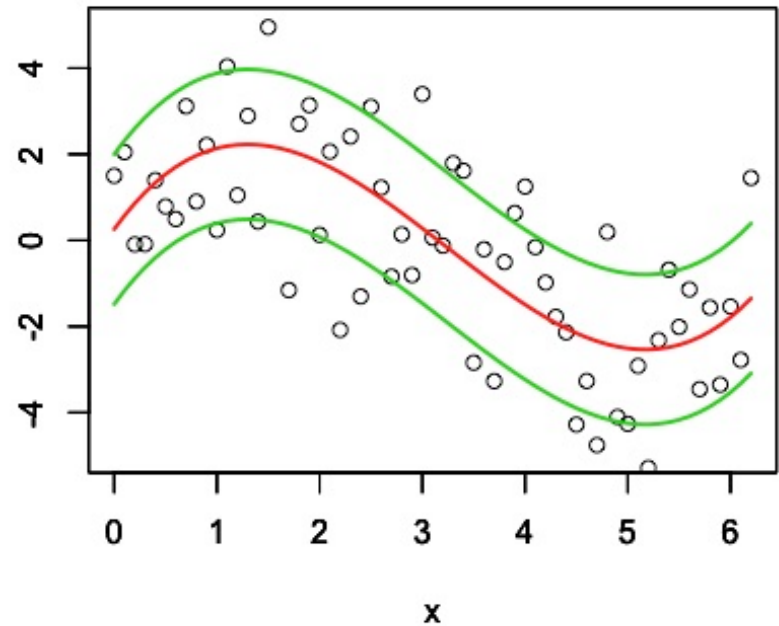
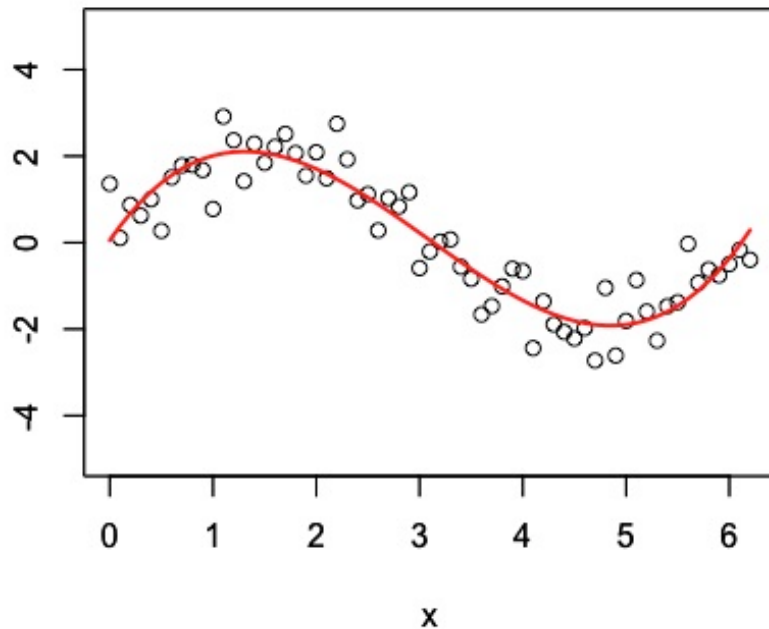
これを最小にする(最小二乗法と一致)

さらに, s について最小化することで s を求められる

分散を含んだ予測分布を求めることできた

実際の例

sin関数にノイズを与えて3次で近似



分布も含めて予測可能

確率的に予測することの利点

本日の流れ

- 機械学習の基礎
 - 機械学習とは？
 - 手法の概要
- 少し専門的な数学の説明
 - 線形代数(エクセル使用)
 - 統計学
- 実装方法の概略(エクセル使用)
- 実際の材料への適用例
 - ガラス材料に応用する場合の課題
 - 世界の研究動向
 - 光学ガラスへの適用の実例

線形代数の復習

連立方程式

$$\begin{aligned} 2x + 3y &= -1 \\ x + y &= 0 \end{aligned}$$



$$\begin{pmatrix} 2 & 3 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

行列 × ベクトル = ベクトル

$$A \times z = b$$

このような表記のメリット

- ✓ 短い式で表すことができる
- ✓ 一般化できる(次元をあげられる)
- ✓ 四則演算を定義すると, 解を求められる

例えば・・・

$$(A \times z) / A = b / A$$

$$z = b / A$$

行列の四則演算

和と差

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} (1+5) & (2+6) \\ (3+7) & (4+8) \end{pmatrix}$$

積

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \times \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} (1 \times 5 + 2 \times 7) & (1 \times 6 + 2 \times 8) \\ (3 \times 5 + 4 \times 7) & (3 \times 6 + 4 \times 8) \end{pmatrix}$$

なぜこのような複雑なことをするのか？



実際の計算の規則と合わせるため

かける順序を逆にしたらどうなるか？
2×3行列の場合にはどうなるか？

行列の四則演算2

商 逆数をかけるのと同様のことを行う

例えば

$$5 \div 5 = 5 \times \frac{1}{5} = 1$$

$A \times A^{-1} = 1$ となるような A^{-1} をかける (逆行列)

ここで, $I \equiv 1$ と表記し,

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

二次元の場合,

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ のとき, } A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

$A \times z = b$ の場合,

$$A^{-1} A \times z = A^{-1} \times b \quad \Rightarrow \quad z = A^{-1} \times b$$

連立一次方程式

連立方程式

$$2x + 3y = -1$$

$$x + y = 0$$

手計算で解いてみる

$$\begin{pmatrix} 2 & 3 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

手計算(逆行列を作って計算)

エクセル で解いてみる

行列やベクトルの微分

参考

$$\vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \text{ のスカラー } x \text{ による微分 } \begin{pmatrix} \frac{\partial a_1}{\partial x} \\ \frac{\partial a_2}{\partial x} \\ \vdots \\ \frac{\partial a_n}{\partial x} \end{pmatrix} \text{ を}$$

$$\left(\frac{\partial \vec{a}}{\partial x} \right)_i = \frac{\partial a_i}{\partial x} \text{ と成分のみで表すことがある}$$

逆にスカラー x のベクトル $\vec{a}$ による微分は,

$$\left(\frac{\partial x}{\partial \vec{a}} \right)_i = \frac{\partial x}{\partial a_i}$$

ベクトル $\vec{a}$ のベクトル $\vec{b}$ による微分は,

$$\left(\frac{\partial \vec{b}}{\partial \vec{a}} \right)_i = \frac{\partial a_i}{\partial b_j}$$



行列になっている

行列やベクトルの微分2

参考

$$\frac{\partial(\vec{x}^T \cdot a)}{\partial x} = \frac{\partial(\vec{a}^T \cdot x)}{\partial x} = \vec{a}$$

証明方法
成分を書き出す

$$\frac{\partial(A \cdot B)}{\partial x} = \frac{\partial A}{\partial x} \cdot B + A \frac{\partial B}{\partial x}$$

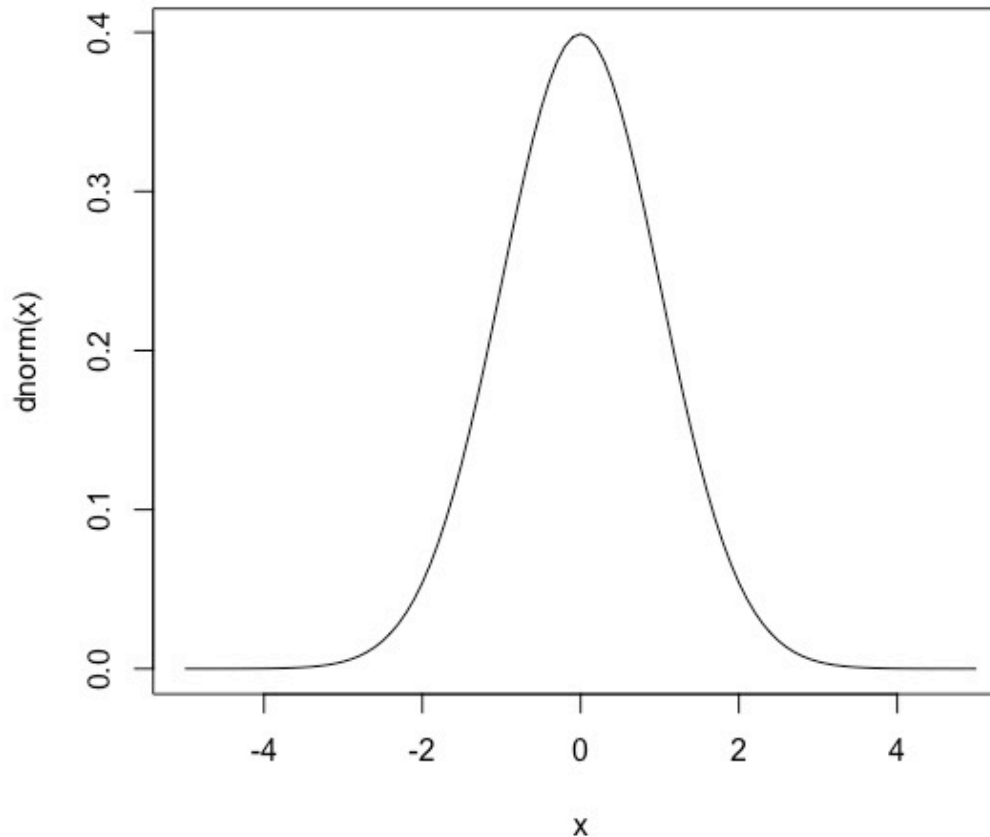
$$\frac{\partial(A^{-1})}{\partial x} = -A^{-1} \frac{\partial A}{\partial x} A^{-1}$$

$A^{-1}A = I$ の微分



正規分布(ガウス分布)

正規分布: 連続変数の分布のモデル



$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right)$$

m: 平均, s: 標準偏差

正規分布の特徴

確率の和 S は

$$S = \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) dx = 1$$

取りうる値 x の平均値は

$$\bar{x} = \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) \times x dx = \mu$$

証明方法
 $y = x - \mu$
と置く

取りうる値 x^2 の平均値は

$$\overline{x^2} = \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) \times x^2 dx = \mu^2 + \sigma^2$$

分散 V の定義より

$$V = \overline{x^2} - \bar{x}^2 = \sigma^2$$

エクセルで計算してみる

D次元のベクトル $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{pmatrix}$ に対して

$$\mathcal{N}(\vec{x}|\vec{\mu}, \Sigma^2) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} (\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu}) \right)$$

ここで $\vec{\mu}$ はD次元の平均ベクトル $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{pmatrix}$

Σ は $D \times D$ の共分散行列



相関の度合いを表す
相関がなければ対角行列

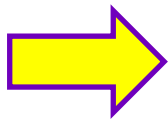
本日の流れ

- 機械学習の基礎
 - 機械学習とは？
 - 手法の概要
- 少し専門的な数学の説明
 - 線形代数(エクセル使用)
 - 統計学
- 実装方法の概略(エクセル使用)
- 実際の材料への適用例
 - ガラス材料に応用する場合の課題
 - 世界の研究動向
 - 光学ガラスへの適用の実例

線形回帰モデル

“モデル”：入力 x に対して出力 y を与えるもの(関数)

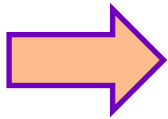
$$y = f(x) \quad \begin{array}{l} f(x) = ax + b \\ f(x) = A \sin x \\ \vdots \end{array}$$



2つの事物の関係性を表現するもの

“学習”：与えられた入力 x と出力 y に合うようにモデル(関数)の形を変えること

$f(x) = \textcircled{a}x + \textcircled{b}$ ここ(パラメータ)がデータ更新のたびに変わる

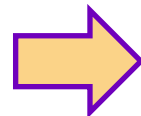


2つの事物の結ぶ“機械”の形を、
新たなデータに合わせて変えていく

入力 x から出力 y を予測する

出力 y が離散的

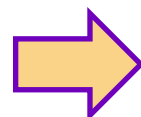
Ex. 身長から性別を予測



分類・識別

出力 y が連続的

Ex. 身長から体重を予測



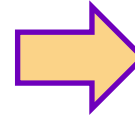
回帰

回帰

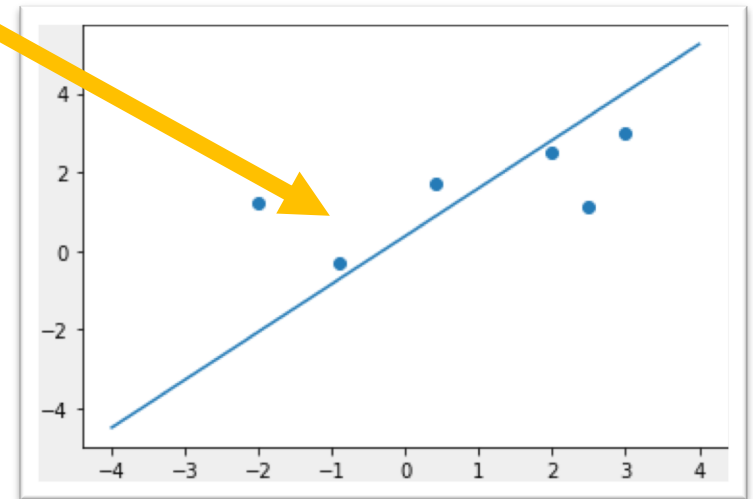
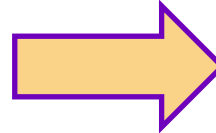
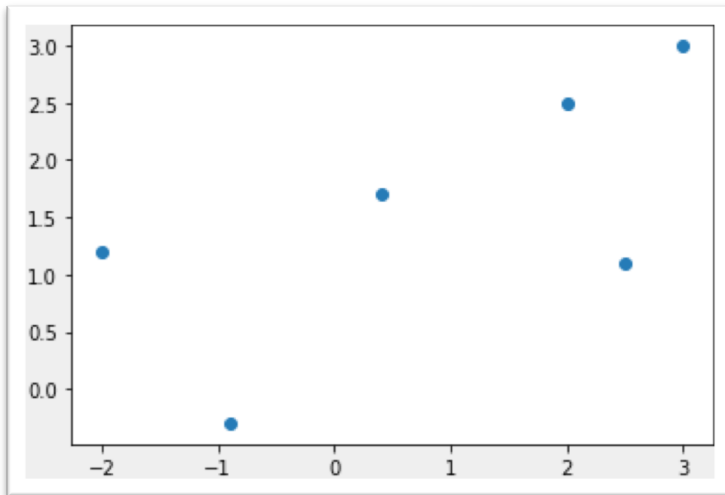
入力 x から出力 y を予測する

入力となる変数 x が1つ

$$y^* = ax + b$$



単回帰



どのようにして直線式を決めるのか？

$$y^* = ax + b$$

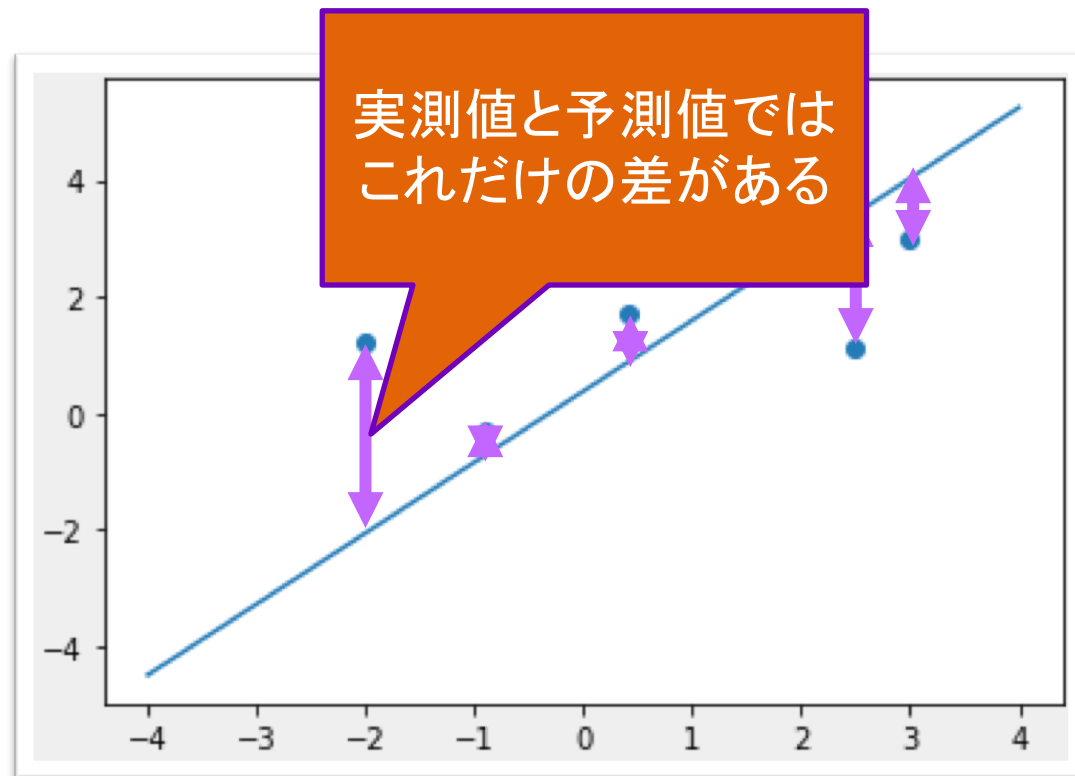
(y^* は予測値)

どのようにして a, b を
求めるか



残差： $y - y^*$

分散： $(y - y^*)^2$

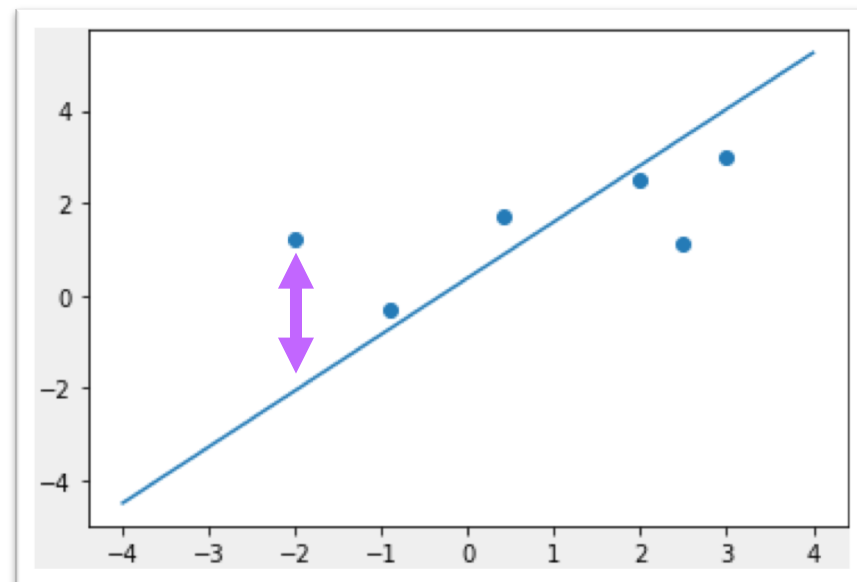


最小二乗法

全てのデータに対し残差が小さい a, b

➡ 残差平方和

$$\begin{aligned}
 E &= (y - y^*)^2 \\
 &+ (y_1 - y_1^*)^2 \\
 &+ \dots \\
 &+ (y_6 - y_6^*)^2 \\
 &= \sum_{n=1}^6 (y_n - y_n^*)^2
 \end{aligned}$$



全データの残差の和が一番小さくなるよう a, b を設定

最小二乗法

$$\begin{aligned}
 E &= \sum_{n=1}^N (y_n - y_n^*)^2 \\
 &= \sum_{n=1}^N \{y_n - (ax_n + b)\}^2 \\
 &= y_n^2 - a^2 x_n^2 + b^2 - 2by_n + 2abx_n - 2ax_n y_n
 \end{aligned}$$

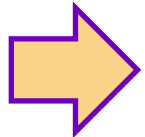
Eが最小になるa, b $\rightarrow$ a, bで偏微分して0

$$\begin{aligned}
 \frac{\partial E}{\partial a} &= \sum_{n=1}^N (2ax_n^2 + 2bx_n - 2x_n y_n) = 0 \\
 \frac{\partial E}{\partial b} &= \sum_{n=1}^N (2b - 2y_n + 2ax_n) = 0
 \end{aligned}$$

最小二乘法

$$\frac{\partial E}{\partial a} = \sum_{n=1}^N (2ax_n^2 + 2bx_n - 2x_n y_n) = 0$$

$$\frac{\partial E}{\partial b} = \sum_{n=1}^N (2b - 2y_n + 2ax_n) = 0$$


$$\begin{aligned} a \sum_n x_n^2 + b \sum_n x_n &= \sum_n x_n y_n \\ a \sum_n x_n + bN &= \sum_n y_n \end{aligned}$$

$$b = \frac{1}{N} (\sum_n y_n - a \sum_n x_n)$$

最小二乗法

$$a \sum_n x_n^2 + b \sum_n x_n = \sum_n x_n y_n \quad \text{について}$$

$$b = \frac{1}{N} (\sum_n y_n - a \sum_n x_n) \text{を代入すると}$$

$$a \sum_n x_n^2 + \frac{1}{N} \sum_n x_n (\sum_n y_n - a \sum_n x_n) = \sum_n x_n y_n$$

$$a (\sum_n x_n^2 - \frac{1}{N} \sum_n x_n \sum_n x_n) = \sum_n x_n y_n - \frac{1}{N} \sum_n x_n \sum_n y_n$$

ゆえに

$$a = \frac{\sum_n x_n y_n - \frac{1}{N} \sum_n x_n \sum_n y_n}{\sum_n x_n^2 - \frac{1}{N} \sum_n x_n \sum_n x_n}$$

$$b = \frac{\sum_n x_n y_n - \frac{1}{N} \sum_n x_n \sum_n y_n}{\sum_n x_n^2 - \frac{1}{N} (\sum_n x_n)^2}$$

前ページの式

$$\begin{aligned} a \sum_n x_n^2 + b \sum_n x_n &= \sum_n x_n y_n \\ a \sum_n x_n + bN &= \sum_n y_n \end{aligned}$$

$$\Rightarrow \begin{pmatrix} \sum_n x_n^2 & \sum_n x_n \\ \sum_n x_n & N \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_n x_n y_n \\ \sum_n y_n \end{pmatrix}$$

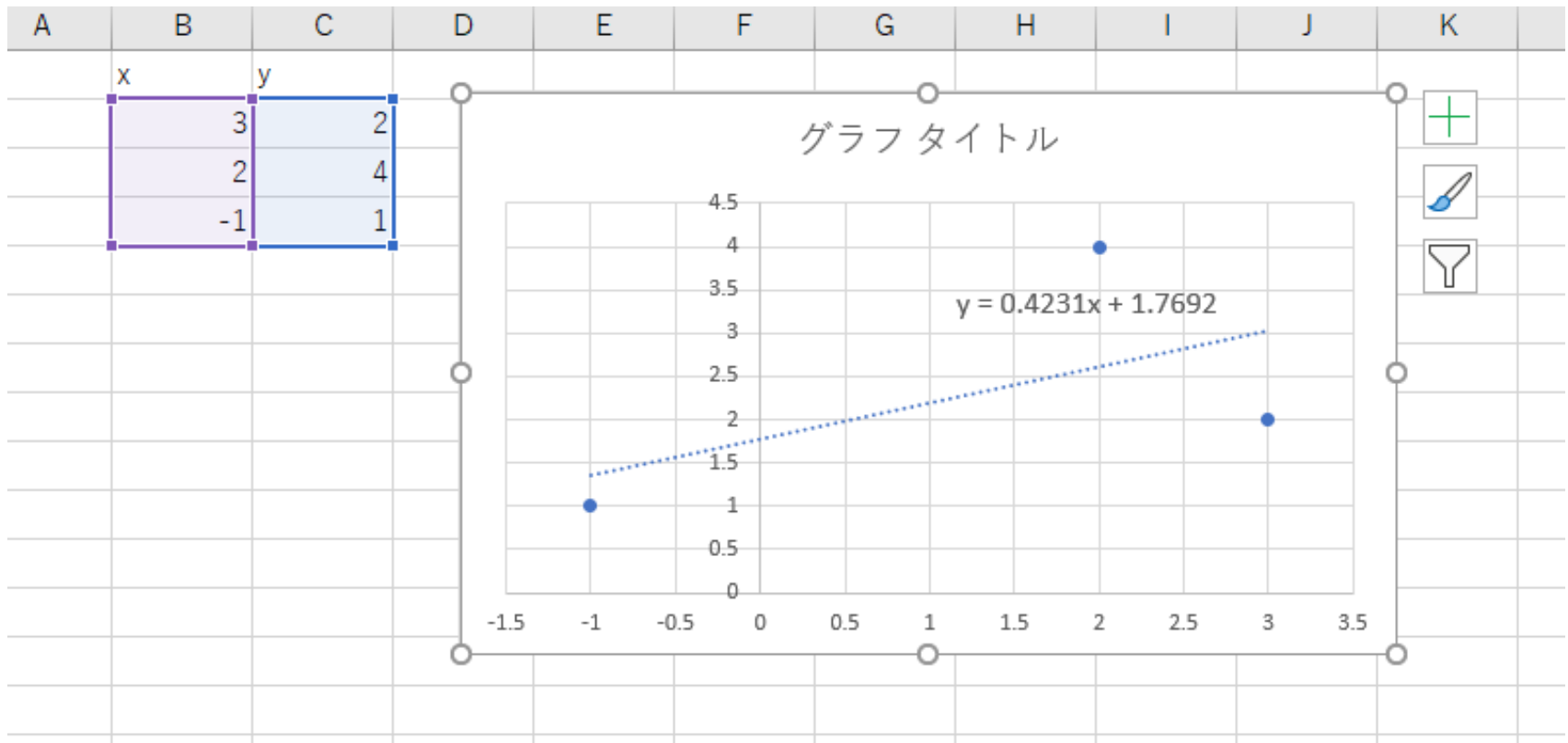
正規方程式

回帰式の導出

ex. $(3,2), (2,4), (-1,1)$ の3点に最もフィットする一次式

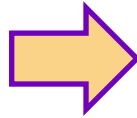
手計算でやってみる
(エクセル使用可)

エクセルによる回帰式の導出



説明変数 x が複数の場合

入力 x が1つ

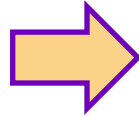


単回帰

$$y^* = ax_1 + b$$

$$x = x_1$$

入力 x が2つ以上



重回帰

$$y^* = ax_1 + bx_2 + cx_3 + \dots \quad x = (x_1, x_2, \dots)$$

どのようにして直線式を決めるのか？

$$y_1^* = w_0 + w_1x_{11} + w_2x_{12} + w_3x_{13} + \dots + w_Dx_{1D}$$

行列で書くと

$$= (w_0, w_1, w_2, w_3, \dots, w_D) \begin{pmatrix} x_{11} \\ x_{12} \\ x_{13} \\ \vdots \\ x_{1D} \end{pmatrix}$$

$$= \boxed{w^T x_1}$$

データが複数あるので...

$$\begin{pmatrix} y^* 1 \\ y^* 2 \\ y^* 3 \\ \vdots \\ y^* D \end{pmatrix} = \begin{bmatrix} 1 & \cdots & x_{1D} \\ \vdots & \ddots & \vdots \\ 1 & \cdots & x_{ND} \end{bmatrix} \begin{bmatrix} w_0 \\ \vdots \\ w_d \end{bmatrix}$$

$$\underbrace{\quad}_{y^*} = \underbrace{\quad}_{\mathbf{X}} \underbrace{\quad}_{\mathbf{w}}$$

計画行列

重回帰での最小二乗法

$$\begin{aligned} E &= \sum_{n=1}^N (y_n - y_n^*)^2 \\ &= \sum_{n=1}^N \{y_n - (w^T x_n)\}^2 \\ &= (y_n - (w^T x_n))^T (y_n - (w^T x_n)) \\ &= y^T y - y^T X w - (X w)^T y + (X w)^T X w \\ &= y^T y - 2w^T (X w)^T + w^T X^T X w \end{aligned}$$

Eが最小になる $w \rightarrow w$ で微分して0となるときの w

色々計算してまとめると

$$\begin{aligned}\frac{\partial E}{\partial \mathbf{w}} &= -2\mathbf{X}^T\mathbf{y} + (\mathbf{X}^T\mathbf{X} + (\mathbf{X}^T\mathbf{X})^T)\mathbf{w} \\ &= -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\mathbf{w} = 0\end{aligned}$$

ゆえに

$$\mathbf{X}^T\mathbf{X}\mathbf{w} = \mathbf{X}^T\mathbf{y}$$

$$\mathbf{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

正規方程式

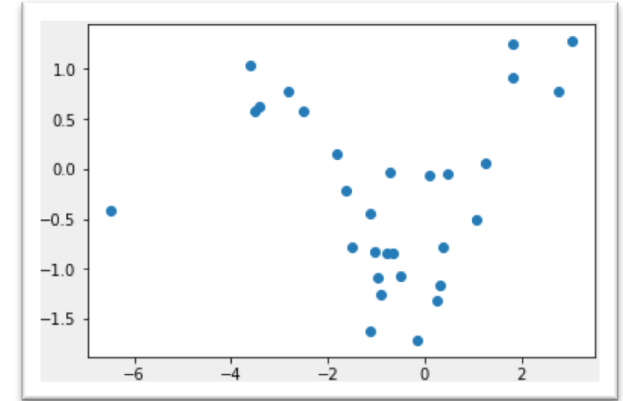
エクセルで挑戦

線形回帰モデル

$$y^* = ax + b$$

直線・平面的

$$y^* = a\Phi_1 + b\Phi_2 + c\Phi_3 + \dots$$



$$y^* = -0.065 + 0.068x + 0.022x^2 + 0.333\sin x - 0.863\cos x$$

$\Phi_1 = x$, $\Phi_2 = x^2$, $\Phi_3 = \sin x$ などのようにして
表現力を上げることができる

$$y^* = w_1\Phi_1 + w_2\Phi_2 + w_3\Phi_3 + \dots$$

Wに関して
線形

基底関数の変え方

$$\underbrace{(y^*_1, y^*_2, \dots, y^*_N)}_{\mathbf{y}^*} = \underbrace{\begin{bmatrix} 1 & \cdots & x_{1D} \\ \vdots & \ddots & \vdots \\ 1 & \cdots & x_{ND} \end{bmatrix}}_{\Phi} \underbrace{\begin{bmatrix} w_0 \\ \vdots \\ w_d \end{bmatrix}}_{\mathbf{w}}$$

$\Phi_1 = x$, $\Phi_2 = x^2$ の場合

$$(y^*_1, y^*_2, \dots, y^*_N) = \begin{bmatrix} 1 & \cdots & x_1^2 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & x_N^2 \end{bmatrix} \begin{bmatrix} w_0 \\ \vdots \\ w_d \end{bmatrix}$$

エクセルで計算してみる

$$y^* = w^T \Phi(x) = w_0 \Phi_0 + w_1 \Phi_1 + w_2 \Phi_2 + w_3 \Phi_3 + \dots + w_H \Phi_H(x)$$

$\Phi = x, x^2, \sin(x), e^x \dots$ など, 無限にある関数からいくつかを選んで重み付き和としている

例えば, $\Phi_0 = 1, \Phi_1 = x, \Phi_2 = x^2$ の場合,

$$y^* = \Phi_0 + 2\Phi_1 + \Phi_2 = 1 + 2x + x^2$$

特徴ベクトル

例えば $\Phi_0 = 1$, $\Phi_1 = x$, $\Phi_2 = x^2$ の場合
 $y^* = \Phi_0 + 2\Phi_1 + \Phi_2 = 1 + 2x + x^2$

$$(y^*_1, y^*_2, \dots, y^*_N) = \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \ddots & \vdots \\ 1 & x_N & x_N^2 \end{bmatrix} \begin{bmatrix} w_0 \\ \vdots \\ w_d \end{bmatrix}$$

$$x \xrightarrow{\text{変換}} (\Phi_1, \Phi_2, \dots, \Phi_d) = (1, x, \dots, \cos(x))$$

変換

特徴ベクトル

先ほどの方法でうまく回帰できない場合

逆行列がない

$x_1 = (2, 4)$, $x_2 = (3, 6)$, $x_3 = (4, 8)$ の場合,

$y^* = w_0 + w_1x_1 + w_2x_2$ に回帰するとき

$$\begin{array}{c} y_1 \\ y_2 \\ y_3 \end{array} = \begin{array}{|c|c|c|} \hline 1 & 2 & 4 \\ \hline 1 & 3 & 6 \\ \hline 1 & 4 & 8 \\ \hline \end{array} \cdot \begin{array}{c} w_1 \\ w_2 \\ w_3 \end{array}$$

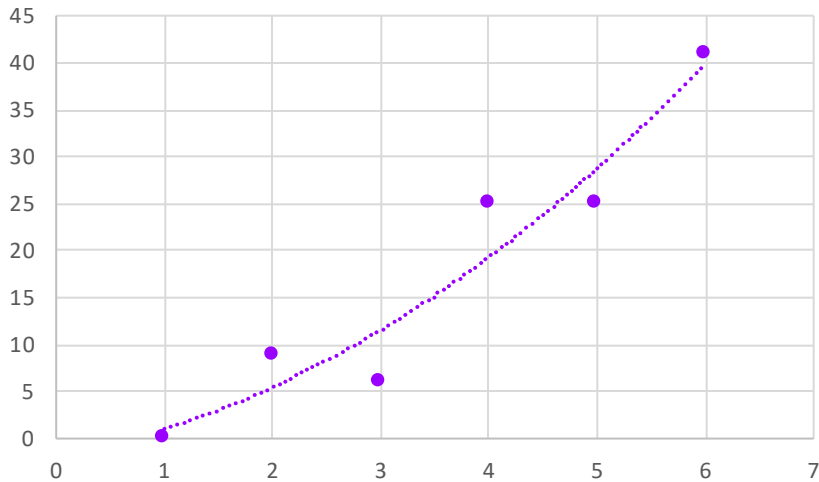
X
 w

逆行列がない
 $w = (X^T X)^{-1} X^T y$

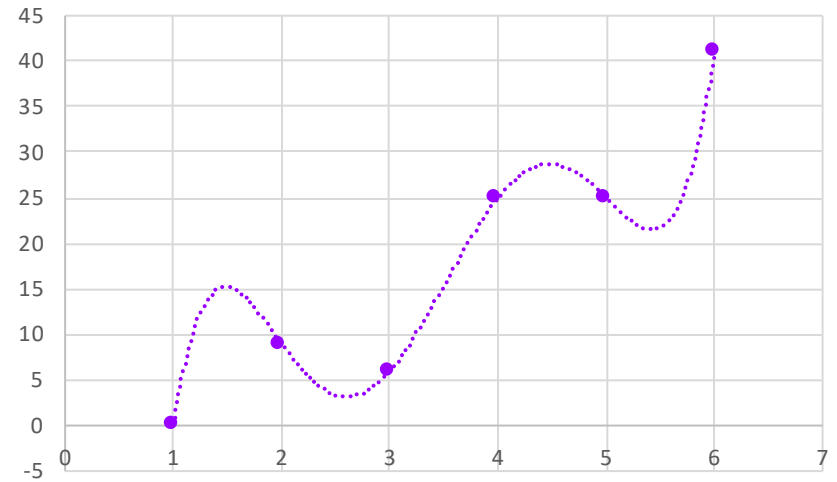
過学習してしまう

過学習とは

2次で近似



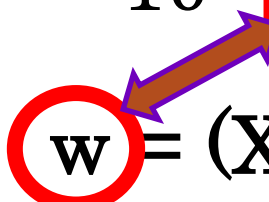
6次で近似



エクセルで計算してみる

リッジ回帰

$$\mathbf{X} = \begin{matrix} & 1 & 2 & 4 \\ \begin{matrix} x = \\ \\ \end{matrix} & 1 & 3 & 6.1 \\ & 1 & 4 & 7.9 \end{matrix} \quad (\mathbf{X}^T \mathbf{X})^{-1} = \begin{matrix} & 6.33 & 18 & -10 \\ & 18 & \boxed{254} & \boxed{-130} \\ & -10 & \boxed{-130} & 66.67 \end{matrix}$$


 $\mathbf{w} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$

➡ Wの絶対値が大きくなりすぎないようにする

$$E = \sum_{n=1}^N (y_n - y_n^*)^2$$

$$\boxed{E} = (y_n - Xw)^2 + \boxed{\alpha w^T w}$$

Wが大きくなると
誤差Eも大きくなる

リッジ回帰の計算方法

$$E = (y_n - Xw)^2 + \alpha w^T w$$

$$\frac{\partial E}{\partial w} = -2X^T y + 2X^T X w + 2\alpha w = 0$$

$$(X^T X + \alpha I) w = X^T y$$

$$w = (X^T X + \alpha I)^{-1} X^T y \quad (\alpha = 0 \text{ の場合最小二乗法と同じ})$$

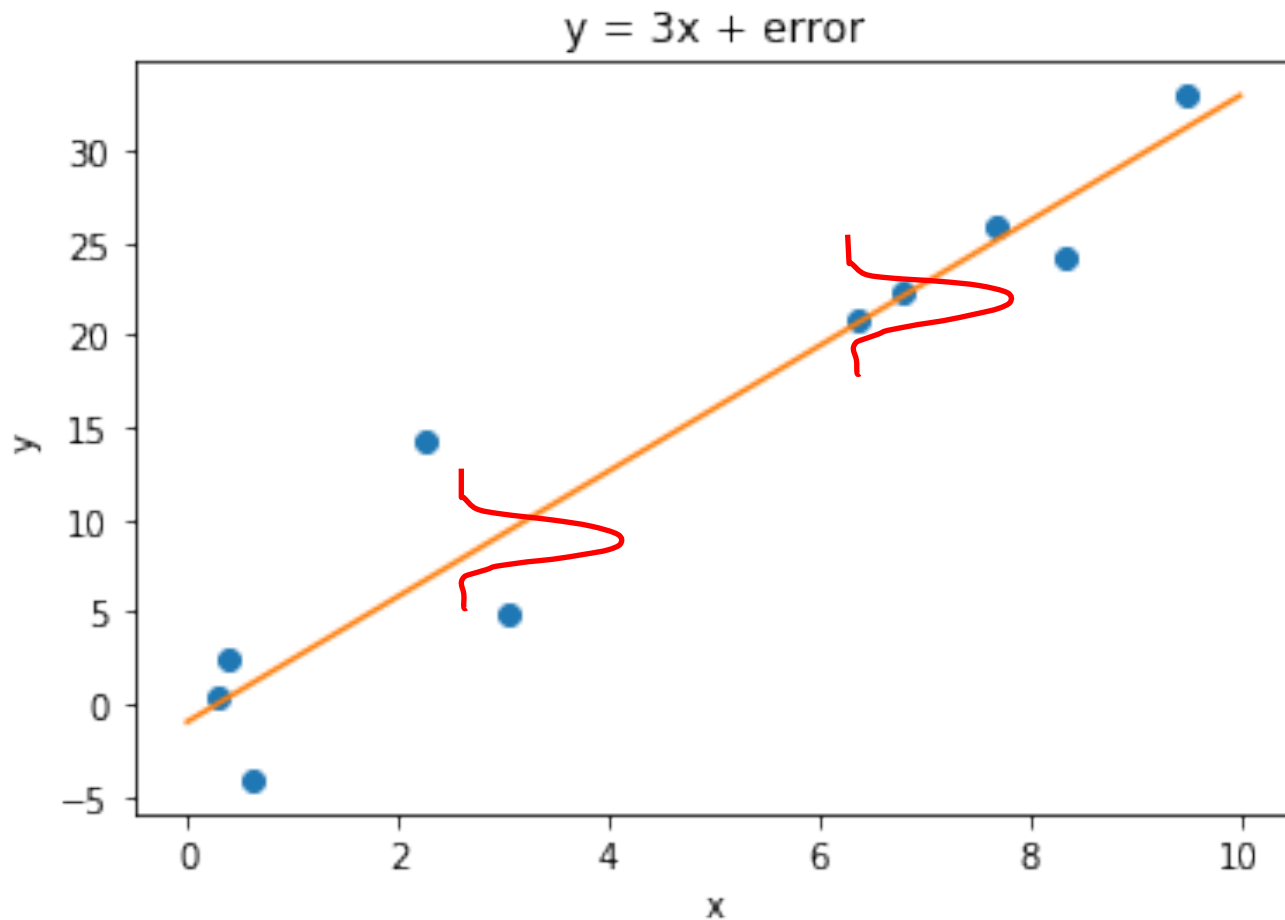
ハイパーパラメータ

正則化項を入れた重み計算で回帰

リッジ回帰

エクセルで計算してみる

確率的な考え方の導入



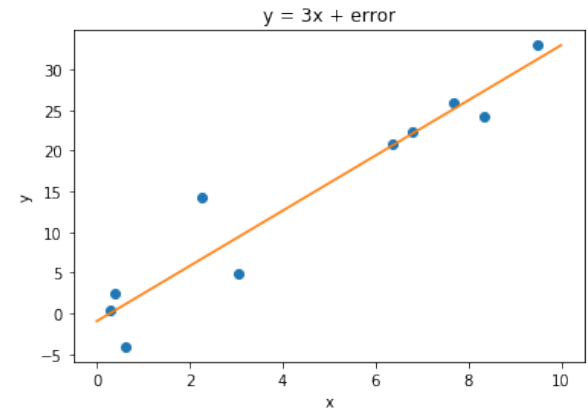
実測のデータは、図のy軸に沿った線上に
ガウス分布している(と考える)

問題の一般化(少し)

与えられているデータ点(訓練データ)

$$(x_1, t_1), (x_2, t_2) \cdots (x_n, t_n)$$

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \rightarrow \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{pmatrix}$$



求めたい回帰曲線

$$y = w_0 + w_1 \cdot x = (w_0 \quad w_1) \begin{pmatrix} 1 \\ x \end{pmatrix} = \vec{w}^T \cdot \vec{x} = y(\vec{w}, \vec{x})$$

このように表記しておく利点

- ・多次元化(x^2, x^3 など), 多成分化(重回帰)が容易
- ・カーネル法への拡張が可能

尤度

ある値 x に対する t の値は、平均値 $y(\vec{w}, \vec{x})$ のガウス分布に従う

その確率は

$$p(t|x, w, \sigma^2) = \mathcal{N}(t|y(\vec{w}, \vec{x}), \sigma^2)$$

と表せる。



訓練データ(この例では10点の測定点)の**発生確率**は

$p(t_1|x_1, w, \sigma^2) = \mathcal{N}(t_1|y(\vec{w}, x_1), \sigma^2)$ などの積で表せる

$$p(\vec{t}|\vec{x}, \vec{w}, \sigma^2) = \prod_{k=1}^n \mathcal{N}(t_k|y(\vec{w}, x_k), \sigma^2) \quad \leftarrow \text{尤度関数}$$

訓練データの発生確率が求まった

最尤法

最尤推定: 訓練データが得られる確率が**最も高く**なるように $\vec{w}$ を求める

尤度関数は $\exp(-x^2)$ の形をしているため、小さな値となりやすく、扱いに注意が必要

$$p(\vec{t}|\vec{x}, \vec{w}, \sigma^2) = \prod_1^n \mathcal{N}(t_k|y(\vec{w}, x_k), \sigma^2)$$

↓ 対数をとる

$$\ln p = -\frac{1}{2\sigma^2} \underbrace{\sum_1^n \{y(\vec{w}, x) - t_n\}^2}_{\text{これを最小にする(最小二乗法と一致)}} - \frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln(2\pi)$$

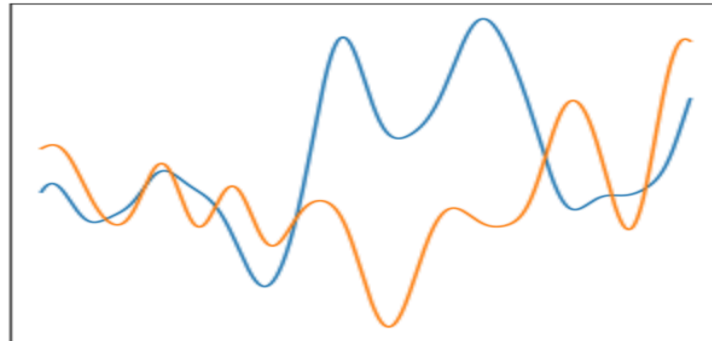
これを最小にする(最小二乗法と一致)

さらに, s について最小化することで s を求められる

分散を含んだ予測分布を求めることできた

ガウス過程回帰

ガウス分布に従う関数でフィット



- 無限次元の回帰(複雑な形状に適用可能)
- ノンパラメトリックな回帰
- カーネルを定義すれば様々な拡張可能

カーネル(グラム行列)

$$K = \begin{pmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{pmatrix}$$



距離を表している
 $\exp(-(x_n - x_1)^2)$

エクセルで計算してみる

グラム行列を定義

$$K = \begin{pmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{pmatrix}$$

回帰関数

$$y = \theta_0 + \alpha_1 \cdot k(x_1, x) + \cdots + \alpha_n \cdot k(x_n, x)$$

線形方程式

$$\alpha = (K + \lambda I)^{-1} Y$$

予測値

$$\text{exp_}Y = K \cdot \alpha$$

エクセルで計算してみる

- 機械学習
 - 「パターン認識と機械学習 上」 C.M. ビショップ (丸善出版)
 - 「カーネル多変量解析」 赤穂 昭太郎 (岩波出版)
 - 「ベイズ推論による機械学習入門」 須山 敦志 (KS情報科学専門書)
 - 「フリーライブラリで学ぶ機械学習入門」 堅田洋資 (秀和システム)
 - Gaussian Processes for Machine Learning, C. E. Rasmussen (The MIT press) (ウェブ版 <http://www.gaussianprocess.org>)
- 数学的な基礎
 - 「統計のための行列代数 上」 D. A. ハーヴィル (丸善出版)
- 実装
 - 「Rによる機械学習入門」 金森敬文 (オーム社)