

---

# **AISECjp 第 5 回 資料**

リリース **2016-08-24**

**Ichiro Fujii**

**9 月 21, 2016**



# 目次

第 3 章 最尤推定法 Maximum Likelihood Estimation	3
3.1 確率モデルの利用	3
3.2 単純化した例による解説	10
参考 URL :	11
AISECjp (Artificial Intelligence x Security) #5 議事録	13
開催情報	13
1. 勉強会概要	13
2. LT : 通信ログ ダッシュボードの試作	13
3. 輪読 : 「IT エンジニアのための機械学習入門」 第 3 章 最尤推定法	14
4. 追加 LT: AI 技術動向の調査結果について	16
Indices and tables	19



- 日時：2016 年 08 月 24 日（水） 19:30～21:00
- 場所：マネーフォワード株式会社

Contents:



## 第3章 最尤推定法 Maximum Likelihood Estimation

- 日時：2016年08月24日（水）19:30～21:00
- 場所：マナーフォワード株式会社
- 発表者：藤井 一郎
- 内容: 最尤推定法を用いた回帰分析

### 3.1 確率モデルの利用

最尤推定法： あるデータが得られる確率が最大となるようなパラメータを推定

- (1) パラメータを含むモデル（数式）を設定する
- (2) パラメータを評価する基準を定める
- (3) 最良の評価を与えるパラメータを決定する

#### 3.1.1 「データ発生確率」の設定

パラメータを含むモデル（数式）を設定する：

- (1) データの背後には  $M$  次多項式の関係があり、さらに標準偏差  $\sigma$  の誤差が含まれている

$$\begin{aligned} f(x) &= w_0 + w_1x + w_2x^2 + \cdots + w_Mx^M \\ &= \sum_{m=0}^M w_mx^m \end{aligned}$$

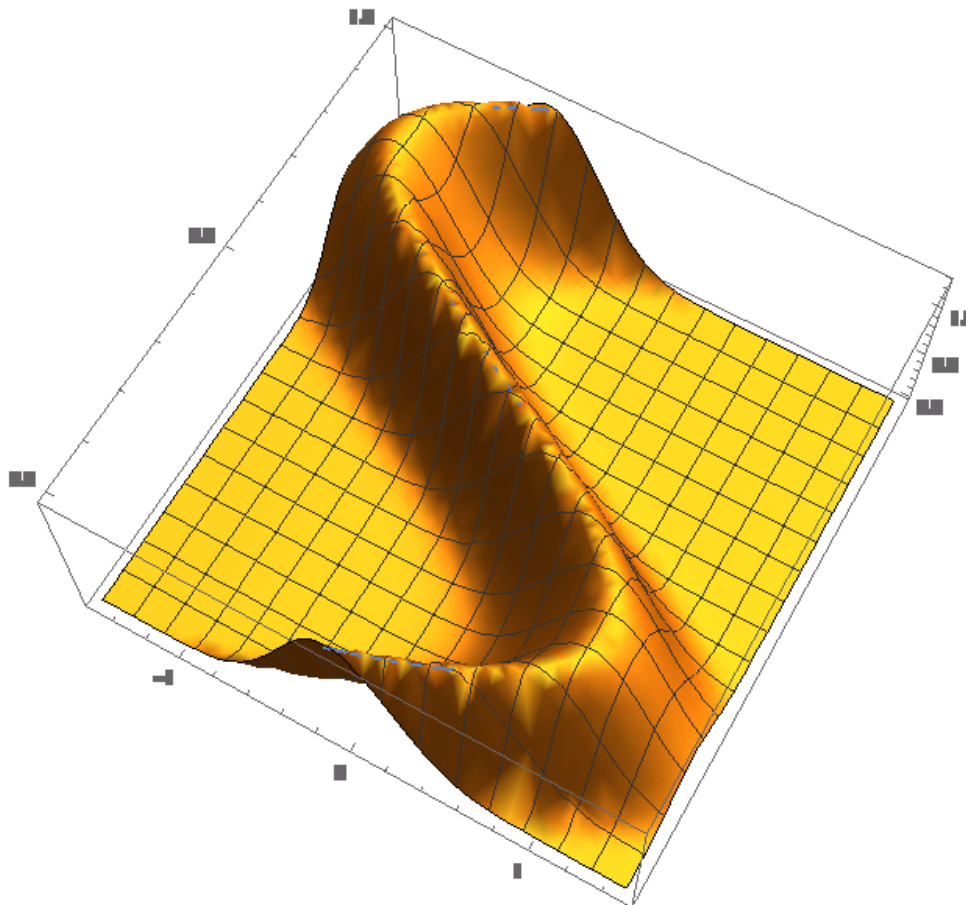
- (2) 観測点  $x_n$  における観測値  $t$  は、 $f(x_n)$  を中心としておよそ  $f(x_n) \pm \sigma$  の範囲に散らばる

(平均  $f(x_n)$ 、分散  $\sigma$  の正規分布)

$$N(t|f(x_n), \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(t-f(x_n))^2}$$

この 2 つの式の  $w_m$  と  $\sigma$  とを推定する。

最小二乗法との違い：データに含まれる誤差を合わせて推定する



### 3.1.2 尤度関数によるパラメーターの評価

パラメータを評価する基準を定める

評価方法 = 尤度関数： トレーニングセットに含まれるデータ  $\{(x_n, t_n)\}_{n=1}^N$  が得られる確率 (パラメータは  $w_m$  と  $\sigma$ )

$$\begin{aligned} P &= N(t_1|f(x_1), \sigma^2) \times \cdots \times N(t_N|f(x_N), \sigma^2) \\ &= \prod_{n=1}^N N(t_n|f(x_n), \sigma^2) \end{aligned}$$



「最尤推定法」

- 「観測されたデータ（トレーニングセット）は、最も発生確率が高いデータに違いない」との仮説
- 確率  $P$  が最大になるようなパラメータを推定
- 尤度関数の最大値問題

$$P = \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(t_n - f(x_n))^2}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N}{2}} \text{Exp} \left[ -\frac{1}{2\sigma^2} \sum_{n=1}^N \{t_n - f(x_n)\}^2 \right]$$

ここで 自乗誤差  $E_p = \frac{1}{2} \sum_{n=1}^N \{f(x_n) - t_n\}^2$

$$P = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N}{2}} e^{-\frac{1}{2\sigma^2} E_p}$$

ここで  $\beta = \frac{1}{\sigma^2}$  とし、 $E_p$  とパラメータ  $w$  の依存関係を明示

$$P(\beta, w) = \left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}} e^{-\beta E_p(w)}$$

これを最大にするパラメータ  $(\beta, w)$  を求める。

この尤度関数<sup>\*1</sup>は 1 より小さい値を取る関数を多数かけ合わせたものなので取扱が面倒。そこで対数を取った対数尤度関数を最大化することとする。<sup>\*2</sup>

$$\ln P(\beta, w) = \frac{N}{2} \ln \beta - \frac{N}{2} \ln 2\pi - \beta E_p(w)$$

対数尤度関数を最大化する条件：

$$\frac{\partial(\ln P)}{\partial w_m} = 0 \quad (m = 0, \dots, M)$$

$$\frac{\partial(\ln P)}{\partial \beta} = 0$$

$w_m$  について：

$$\frac{\partial E_p}{\partial w_m} = 0 \quad (m = 0, \dots, M)$$

これは自乗誤差を最小にする条件と同じ：

<sup>\*1</sup> 当初の記述は「単調増加関数なので対数をとっても単調増加する。（対数尤度関数）」としており、勉強会終了後懇親会で尤度関数は単調増加ではないのでは？との質問があった。

<sup>\*2</sup> 説明資料公開後、尤度関数は単調増加ではないと参考 URL(4.) とともに再度ご指摘いただく。確かにその通りでした。m(\_ \_)m

多項式の係数  $\{w_m\}_{m=0}^M$  は最小二乗法と同じ

$$\sum_{n=1}^N \left( \sum_{m'=0}^M w_{m'} x_n^{m'} - t_n \right) x_n^m = 0$$

$$\sum_{m'=0}^M w_{m'} \sum_{n=1}^N x_n^{m'} x_n^m - \sum_{n=1}^N t_n x_n^m = 0$$

$$\Phi = \begin{pmatrix} x_1^0 & x_1^1 & \cdots & x_1^M \\ x_2^0 & x_2^1 & \cdots & x_2^M \\ \vdots & \vdots & \ddots & \vdots \\ x_N^0 & x_N^1 & \cdots & x_N^M \end{pmatrix}$$

$$w^T \Phi^T \Phi - t^T \Phi = 0$$

$$w = (\Phi^T \Phi)^{-1} \Phi^T t$$

$\beta$  について :

$$\frac{1}{\beta} = \frac{2E_p}{N}$$

$$\sigma = \sqrt{\frac{1}{\beta}} = \sqrt{\frac{2E_p}{N}} = E_{RMS}$$

$$= \sqrt{\frac{1}{N} \sum_{n=1}^N \left( \sum_{m=0}^M w_m x_n^m - t_n \right)^2}$$

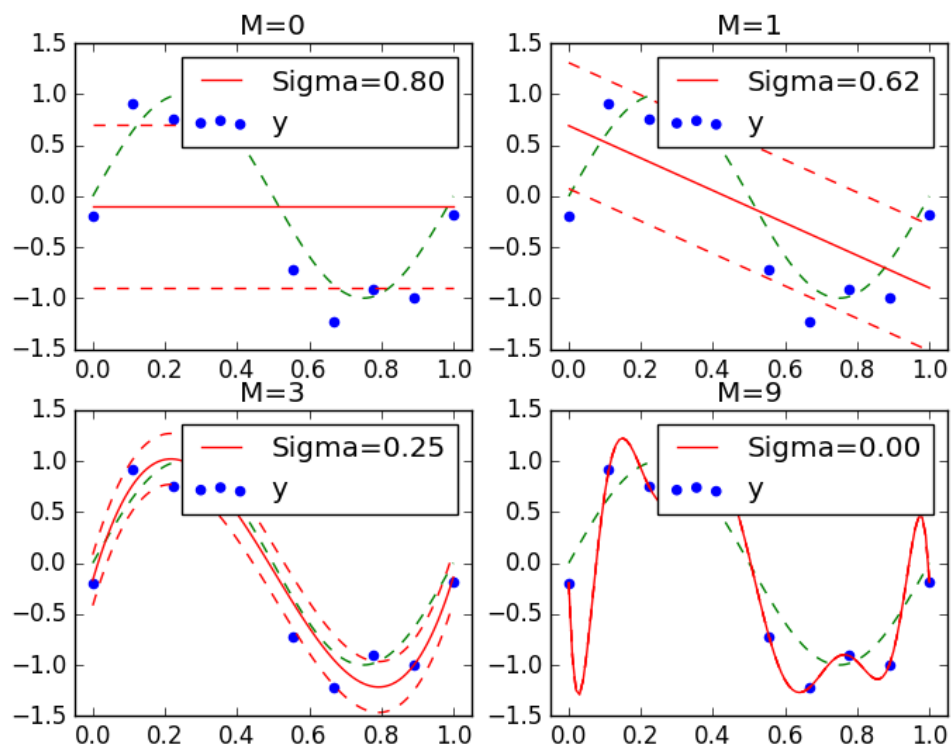
これは最小二乗法の平方根平均自乗誤差

最小二乗法とは異なるアプローチで計算したが、得られた多項式は同じ

最小二乗法は最尤推定法の中でも正規分布の誤差を仮定した特別な場合

### 3.1.3 サンプルコードによる確認

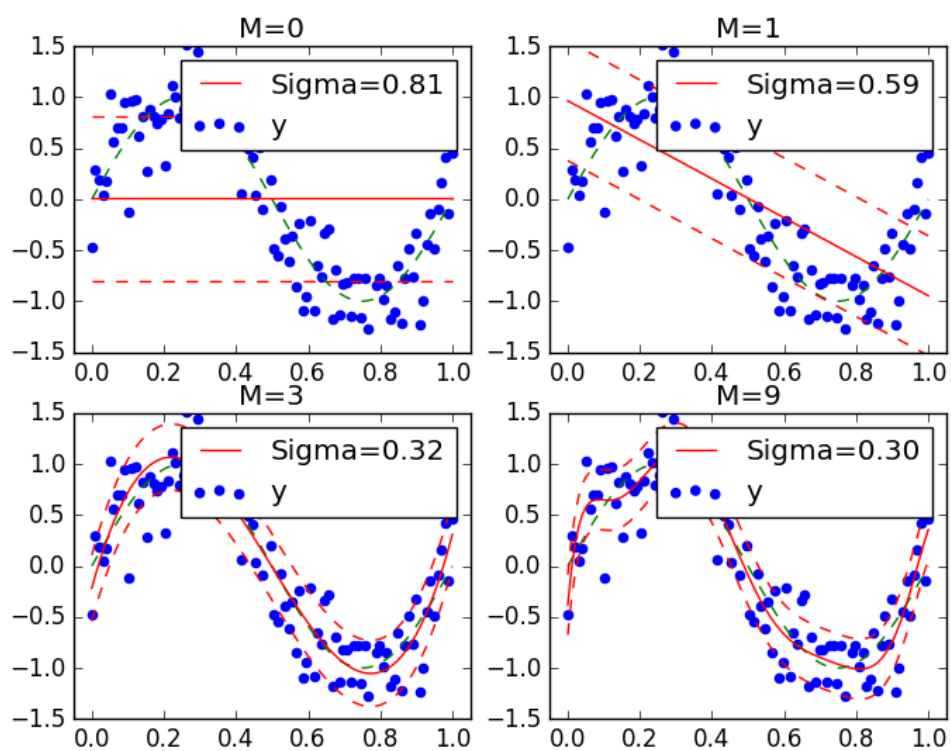
計算結果 :



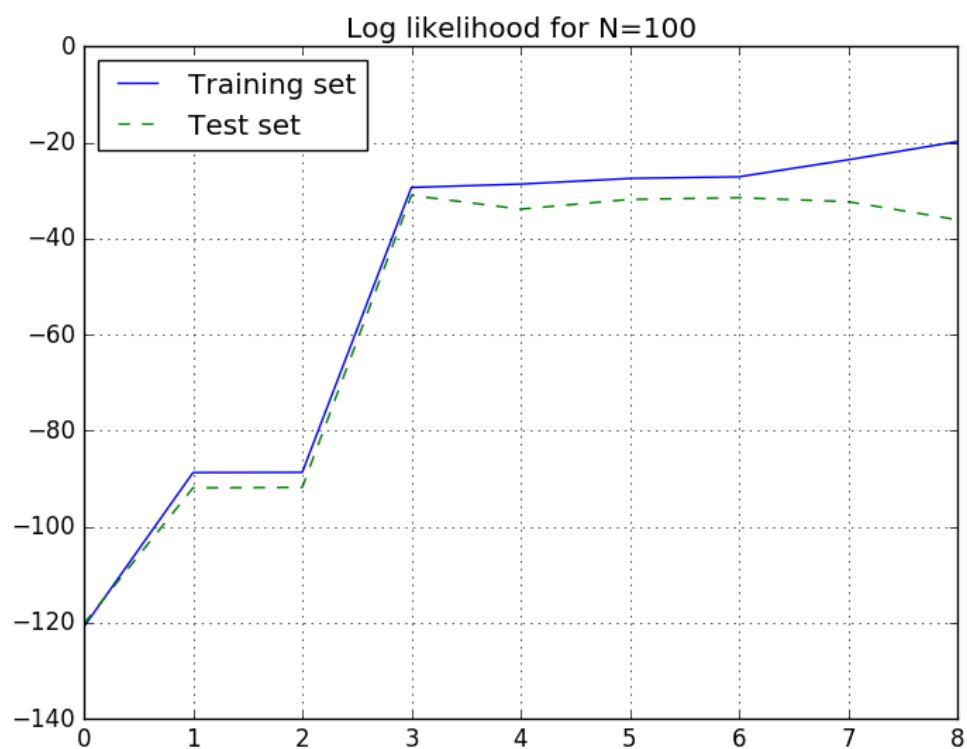
(1)  $N=10$  の結果

対数尤度の変化を見ることでオーバーフィッティングを調べることができる





(2)  $N=100$  の結果



### 3.2 単純化した例による解説

3.1 節では複数の観測点における観測値の予測を行った。この節では、ある観測点に固定して、繰り返し観測値を取得したデータから平均  $\mu$ 、標準偏差  $\sigma$  を最尤推定法で推定してみる。

$$\mu = \frac{1}{N} \sum_{n=1}^N t_n \quad (\text{標本平均})$$

$$\sigma^2 = \frac{1}{N} \sum_{n=1}^N (t_n - \mu)^2 \quad (\text{標本分散})$$



推定値（標本分散）は実際の値（母分散）よりも小さくなる傾向がある。（偏りがある）

偏りをなくするために推定値より大きくしてやる（不偏推定量）

$$\sigma^2 = \frac{1}{N-1} \sum_{n=1}^N (t_n - \mu)^2 \quad (\text{不偏分散})$$

なぜ、 $N$  ではなく  $N-1$  で割るか？

計算式の中に標本平均が含まれているので、 $(N-1)$  個の観測データがあれば他のひとつの観測データは正確に値が決められる状態になる。（自由度が  $N-1$ ）

（厳密な証明は割愛）

## 参考 URL :

- 1. 人工知能に関する断創録 最尤推定、MAP 推定、ベイズ推定 <http://cp.the-premium.jp/>

- 2. 最尤法によるパラメータ推定の意味と具体例 I 高校数学の美しい物語 <http://mathtrain.jp/mle>
- 3. (おまけ) イラストでわかる自由度と不偏分散 <http://home.a02.itscom.net/coffee/tako08Annex2.html>
- 4. やる夫で学ぶ機械学習 - 対数尤度関数 - <http://tkengo.github.io/blog/2016/06/16/yaruo-machine-learning6/>



# AISECjp (Artificial Intelligence x Security)

## #5 議事録

### 開催情報

- エンジニアのエンジニアによるエンジニアのための機械学習勉強会
- 日時：2016 年 8 月 24 日 19:30～21:30
- 場所：マネーフォワード株式会社

### 1. 勉強会概要

\* 機械学習を各自の専門分野の実務で活用することを目的に、まずは基礎理論の解説本の輪読を行う。 \* 各自の取り組み状況についての LT もあり。 \* 開催間隔：月 1 回

### 2. LT：通信ログ ダッシュボードの試作

#### 発表者：

- Suzuki Kengo 氏
- マネーフォワード さん
- セキュリティエンジニア（防御側）
- 本勉強会の発起人の 1 人

内容 :

通信ログを可視化, 異常検知した結果を表示するダッシュボードの試作を始めた。

オープンソースの grayLog を利用。(Suplunk は高くて実用には難あるため)

- <https://www.graylog.org/>
- データ蓄積・分析のバックエンドは mongoDB+ElasticSearch が使われている。
- ElasticSearch の可視化によく使われている Kibana よりはシンプルで使いやすそう。(Kibana は使ったことはないが)
- 各種ログ形式を標準装備, マーケットプレイスから追加のログ形式や追加機能プラグインを取り込むことも可
- 全ての機能が無償で利用できる。サポートが必要な場合はサブスクリプション契約。

構築中のシステムの現状はアラート出まくりのオオカミ少年状態。実用化には適切なアラートに以下に絞り込むかが課題。

ディスカッション :

1. 機器によって形式・内容が異なる等ログのみの解析には限界あるので, 最近のトレンドはパケット解析に移っている。ただし, 膨大な量のバイナリデータを解析する必要があるので人では実質不可能。AI 的手法を活用すべき分野の一つ。
2. PaloAlto では異常検知されなかったがパケット解析により判明した社内インシデント事例の紹介あり。
3. grayLog は日本ではまだほとんど知られていないツール。ユーザ会もない。先駆者になれるかも。

### 3. 輪読 : 「IT エンジニアのための機械学習入門」 第 3 章 最尤推定法

発表者

\* Fujii Ichiro

- aL&IHI エスキューブ
- 重工メーカーの社内業務システムの開発・運用を経て現在は技術企画を担当

## 内容：

本書における機械学習のステップは以下で説明されている。

(1) パラメータを含むモデル（数式）の設定 (2) パラメータを評価する評価基準の設定 (3) 最良の評価を与えるパラメータの決定

最尤推定法を使った回帰分析としては以下の手順となる。

(1)  $N$  次多項式のモデルを設定。パラメータは多項式の係数と、データに含まれる標準偏差データは正規分布に従うと仮定

(2) 評価基準は尤度関数（データの出現確率の掛け合わせたもの）

(3) 尤度関数を最大化する条件：対数尤度関数を多項式係数および標準偏差の偏微分が 0

- 得られる結果の多項式は、最小二乗法で得られる結果と同一。（データが正規分布に従うと仮定したため。）
- 最小二乗法は最尤推定法の特殊ケースともいえる。
- 最尤推定法をさらに拡張したのがベイズ推定法（第 8 章で解説）

## ディスカッション：

(1) この手法はどんな分野に応用できるのか？うまいたとえ話があれば。

- 最尤推定法ではないが、ベイズ推定法は迷惑メールフィルタ、WAF 等で実用されている。

(2) 電子部品メーカーので外観検査への機械学習応用を検討している。自分はソフトウェアエンジニアだが数式に立ち入った理解が必要か、それともツールとして使えればよいのか？

- ツールが出そろっているのですまずは使ってみるというアプローチもあり。ただし、データ活用のためには一般的には、 $\hat{a}$  業務知識,  $\hat{a}$  数学・統計理論,  $\hat{a}$  ソフトウェアエンジニアリングの 3 要素が必要。すべてをまかなえる人はいないのでチームで取り組むのがよい。
- 外観検査ではデータ取る部分でどうしても人手がかかかってしまう。（角度を変えて見てみる等）将来的には検査対象をロボットアームが動かすような仕組みも必要かもしれない。

(3) モデルを 3 次元グラフで可視化したのが分かりやすかった。どんなツールを使ったのか？

- Mathematica という数式処理ソフトを使用。
- <https://www.wolfram.com/mathematica/index.ja.html?footer=lang>
- 数式とプロット範囲を与えればコマンド 1 行でグラフが描ける（プログラム不要）。

- 買えば1本数十万のソフトだが、なぜか 5000 円で買える RaspberryPi にバンドルされている。
- 私はこのために RaspberryPi を買った。

## 4. 追加 LT: AI 技術動向の調査結果について

発表者：

総合 IT ベンダー K2(@bonjinjpn) (フォレンジック技術者)

内容&ディスカッション：

社内の勉強会用に公開情報をまとめた AI 技術動向の調査結果について LT。

機械学習を実務で使うためには、対象の実務知識、数学・統計の知識、プログラミング能力+英語力が必要ですべてを分かるスーパーマンはいない。そこで社内でチームとして取り組めるように勉強会を始めた。

(1) 情報処理学会コンピュータセキュリティ研究会の論文における機械学習関連の論文数は 2014 年から急増している。

(2) 学会のスポンサー枠を見るとどんな企業がその分野に力を入れているのかが分かる。例えば CVPR, ECCV や ICCV などの国際学会ではパナソニックや Disney Research がスポンサー枠を持っている。

※参考 URL

<http://www.eccv2016.org/sponsors/>

<http://pamitc.org/iccv15/>

<http://cvpr2016.thecvf.com/>

(3) 現在第 3 次 AI ブームと言われているが少々過熱気味、また冬の時代が訪れるかも。ただし、これまでの違いは GPU (ハード) とオープンソースのツール (ソフト) を使って個人が趣味で実績上げているので意外と長続きする可能性も。

(4) 人狼知能プロジェクト AI による自動文章生成 それなりに意味のある文章が生成できるようになっている。

(5) ROBORACE 自動運転によるフォーミュラーカーレース

※参考 URL

<http://www.roborace.com/>

(6) 現在 AI の応用分野は画像認識分野が多いが, AI で実績出ているのはごく一部の領域。まだまだ取り組むべき領域あり。(たとえば, 3 次元画像復元等)

## 5. まとめ

今回, 初参加のメンバーがほとんどだったが参加者同士のディスカッションが活発に行えて大変有意義だった。



# Indices and tables

- [genindex](#)
- [modindex](#)
- [search](#)