パターン認識と機械学習
1.2.4 ガウス分布

小野寺喜行

ガウス分布

- 単一の実数値変数xに対し

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

と定義

- 以下の2つのパラメータを持つ

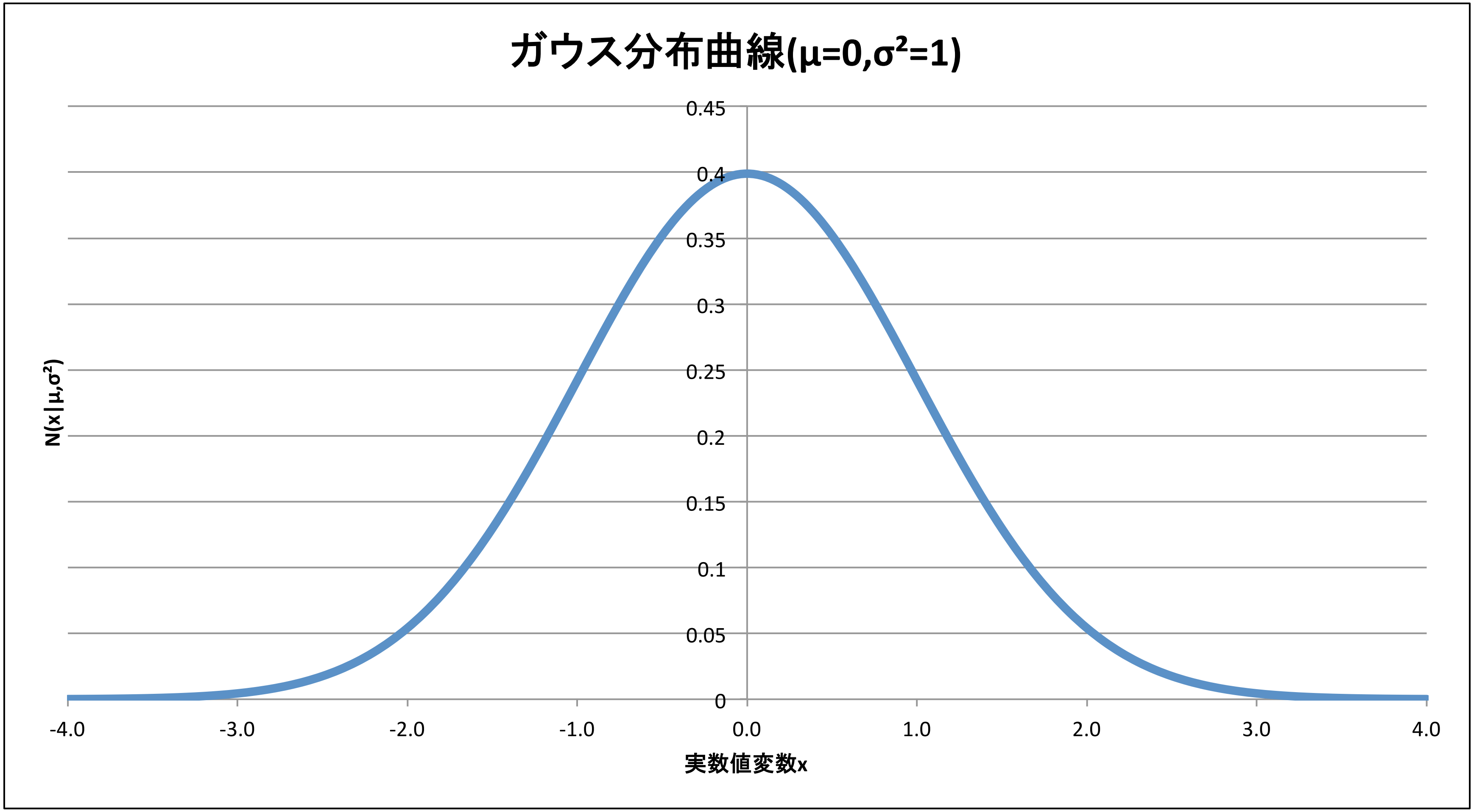
  平均（mean）: μ

  分散(variance）: σ^2

分散

- 分散の平方根 σ :
  標準偏差（standard deviation）
- 分散の逆数 $\beta = 1/\sigma^2$:
  精度パラメータ（presicion parameter）

ガウス分布 曲線



ガウス分布曲線(μ=0,σ²=1)
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
N(x|μ,σ²)
-4.0
-3.0
-2.0
-1.0
0.0
1.0
2.0
3.0
4.0
実数値変数x

確率密度の用件

- 定義より

$$N(x \mid \mu, \sigma^2) > 0$$

- また、

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

- 確率密度の満たすべき2つの用件を満たす

平均

- xの平均値

$$\mathrm{E}\left[x\right] = \int_{-\infty}^{\infty} N(x \mid \mu, \sigma^2) x \, dx = \mu$$

分散

- 2次のモーメント

$$\mathrm{E}\left[x^2\right] = \int_{-\infty}^{\infty} N(x \mid \mu, \sigma^2) x^2 \, dx = \mu^2 + \sigma^2$$

- x の分散

$$\mathrm{var}\left[x\right] = \mathrm{E}\left[x^2\right] - \mathrm{E}\left[x\right]^2 = \sigma^2$$

モード（最頻値）

- 分布の最大値を与えるxはモード（最頻値）である。
- ガウス分布では、モードは平均に一致する。

D次元ベクトル

- D次元ベクトルの連続変数 $\mathbf{x}$ に対してのガウス分布

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

D次元ベクトル$\boldsymbol{\mu}$: 平均

D×D行列Σ : 共分散

$|\boldsymbol{\Sigma}|$:Σの行列式

独立同分布

- データ点が同じ分布から独立に生成されるとき、<u>**独立同分布**（**independent identically distributed**）（**i.d.d.**）</u>であるという。
- 2つの独立な事象の同時確率は、それぞれの事象の周辺確率の積で与えられる。

データ集合x

- スカラー変数 x のN個の観測値からなるデータ集合 $\mathbf{x}=(x_1,\dots,x_N)^{\mathrm{T}}$
- データ集合 $\mathbf{x}$ はi.d.d.であるので、μとσ^2が与えられたとき、データ集合の確率は

$$p(\mathbf{x}\mid\mu,\sigma^2)=\prod_{n=1}^{N}N(x_n\mid\mu,\sigma^2)$$

- μとσ^2 の関数とみなすと、ガウス分布に対する尤度関数である

データ集合からパラメータを決める

- 観測されたデータ集合を使って確率分布のパラメータを決める方法

⇩

尤度関数を最大にするようなパラメータの値を求める

- 実際には尤度関数の対数を最大化する方が便利

尤度関数の対数を最大化

- 対数は単調増加関数であるので、関数そのものの最大化と等価
- その後の数学的な解析を単純化
- 小さな確率値の積を計算する際の、計算機の数値精度のアンダーフローを解決できる

対数尤度関数

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

  $$p(\mathbf{x} \mid \mu, \sigma^2) = \prod_{n=1}^{N} N(x_n \mid \mu, \sigma^2)$$

  より

$$\ln p(\mathbf{x} \mid \mu, \sigma^2) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n - \mu)^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi)$$

最尤推定の解

- μ に関して対数尤度関数を最大化すると、最尤推定の解が得られ、

$$\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

  で与えられる
- サンプル平均（標本平均；sample mean）、すなわち観測値$\{x_n\}$の平均

分散に対する最尤解

- σ^2 に関して対数尤度関数を最大化すると、分散に対する最尤解

$$\sigma_{ML}^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{ML})^2$$

を得る

- サンプル平均 μ_{ML} に関するサンプル分散（標本分散；sample variance）である

同時最大化

- μ とσ^2に関して対数尤度関数を同時最大化を行なうとき、ガウス分布の場合はμ とσ^2は分離して解く事が出来る。

最初にμ_{ML}を評価

$\Downarrow$ その結果を使って

σ^2_{ML}を評価

最尤アプローチの重大な限界

- 1変数ガウス分布の最尤パラメータの設定に関する問題
  ⇨最尤のアプローチでは特に分布の分散が系統的に過小評価されている
- バイアス（bias）とよばれる現象の一例
- 多項式曲線のフィッティングにおける過学習の問題に関連

分散の過小評価

- 最尤解μ_{ML}, σ^2_{ML}はデータ集合の値$x_1, \ldots, x_N$の関数であることに注意
- パラメータμ, σ^2をもつガウス分布に従うデータ集合に関する期待値

$$\mathrm{E}\left[\mu_{ML}\right] = \mu$$

$$\mathrm{E}\left[\sigma^2_{ML}\right] = \left(\frac{N-1}{N}\right)\sigma^2$$

分散が(n-1)/N倍過小評価されることが示される

分散パラメータの不偏推定量

- $\mathrm{E}\left[\sigma_{ML}^2\right] = \left(\frac{N-1}{N}\right)\sigma^2$ より

$$\tilde{\sigma}^2 = \frac{N}{N-1}\sigma_{ML}^2 = \frac{1}{N-1}\sum_{n=1}^{N}(x_n - \mu_{ML})^2$$

は分散パラメータの不偏推定量になる

最尤解のバイアス

- 最尤解のバイアスはデータ点の数Nが増えればあまり重大ではなくなる
- $N \to \infty$ の極限では分散の最尤解は真の分散に一致する

最尤解のバイアス

- N が小さいという理由以外では、深刻な問題にはならない
- しかし、多くのパラメータを持つ複雑なモデルでは、最尤推定に伴う問題ははるかに厳しいものとなる
- 過学習の問題の根本にある