輪講 - 第3回

Ryosuke Horiuchi

Tokyo, Japan

June 18, 2019

Contents

1. Gaussian distribution(**ガウス分布**)
2. **ガウス分布と最尤推定**
3. **曲線フィッティング**
4. **ベイズ曲線フィッティング**

ガウス分布の確率密度関数

ガウス分布は正規分布とも言われ、平均 μ, 分散 σ^2 を用いて、

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

と表せる.
x が D 次元の場合には、平均 μ, 共分散 Σ を用いて

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

ガウス分布と最尤推定

N **回の観測によって得られた観測値を** $\mathbf{x} = (x_1, \ldots, x_N)^{\mathrm{T}}$ **とする．ただし、これらは** i.i.d (independent and identically distrubuted) **と仮定**.
この時、その事象が起こる確率は、

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

となる.
しかし、実際、μ, σ^2 **は未知 – 最尤推定**

- **最尤推定**
  指数関数族に対しては、対数尤度を最大化することが多い.
  解析的に解きやすくなるだけでなく、数値計算の時も overflow **を起こしにくくなる**.

ガウス分布と最尤推定

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

これの μ, σ^2 **による偏微分より、**

$$\mu_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N} x_n = \textbf{標本平均}$$

$$\sigma^2_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu_{\mathrm{ML}})^2 = \textbf{標本分散}$$

となる.

ガウス分布と最尤推定

最尤推定でも求めた平均 μ_{ML} と分散 σ^2_{ML} の期待値は、

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

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

[証明]

$$
\begin{aligned}
\mathbb{E}\left[\sigma^2_{\mathrm{ML}}\right] &= \mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N}\left(x_n - \frac{1}{N}\sum_{k=1}^{N}x_k\right)^2\right] && (1)\\
&= \mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N}\left(x_n^2 - \frac{2}{N}x_n\sum_{k=1}^{N}x_k + \frac{1}{N^2}(\sum_{k=1}^{N}x_k)^2\right)\right] && (2)\\
&= \mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N}x_n^2 - \frac{1}{N^2}(\sum_{k=1}^{N}x_k)^2\right] && (3)\\
&= \mathbb{E}\left[\frac{N-1}{N^2}\sum_{n=1}^{N}x_n^2 - \frac{1}{N^2}\sum_{i\neq j}x_ix_j\right] && (4)\\
&= \frac{N-1}{N^2}\sum_{n=1}^{N}\mathbb{E}\left[x_n^2\right] - \frac{1}{N^2}\sum_{i\neq j}\mathbb{E}\left[x_ix_j\right] && (5)\\
&= \frac{N-1}{N^2}N(\sigma-\mu^2) - \frac{1}{N^2}N(N-1)(-\mu^2) && (6)\\
(\because\ & x_i : \text{i.i.d}, \text{for} i = 1,\ldots,N) && (7)\\
&= \frac{N-1}{N}\sigma^2 && (8)
\end{aligned}
$$

ガウス分布と最尤推定

e.g) **実際、図 3 の時には、分散は平均しても正しいものとはならない**.

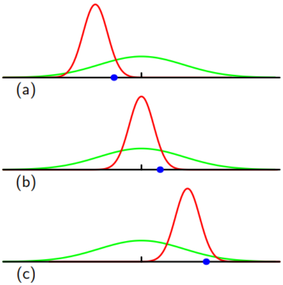

Figure: **バイアスの発生**

つまり、平均はその平均を取ることで真の平均となるが、分散の平均は以前として小さく評価されてしまう. N **が大きい時は問題ないがモデルが複雑になると注意しなければならない**.

曲線フィッティング

観測データが以下のように生成されているというモデルを立てる.

$$p(t|x, \mathbf{w}, \beta) = \mathcal{N}\left(t|y(x, \mathbf{w}), \beta^{-1}\right)$$

ただし、$\beta^{-1} = \sigma^2$ **とする.**

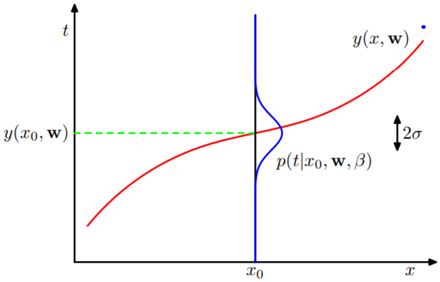

Figure: **観測モデル**

曲線フィッティングと最尤推定

訓練データ $\{\mathrm{x}, \mathrm{t}\}$ から未知パラメータの W, β を最尤推定を用いて求める.

$$p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}\left(t_n | y\left(x_n, \mathbf{w}\right), \beta^{-1}\right)$$

対数尤度を取ると、

$$\ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = -\frac{\beta}{2} \sum_{n=1}^{N} \{y\left(x_n, \mathbf{w}\right) - t_n\}^2 + \frac{N}{2} \ln \beta - \frac{N}{2} \ln(2\pi)$$

つまり、2 乗和誤差を最小化することは、元のデータがあるガウス分布に従ってると仮定した上で対数最尤推定をすることと等価.

曲線フィッティングと最尤推定

最尤推定より求めた $\mathbf{w}_{\mathrm{ML}}, \beta_{\mathrm{ML}}$ を用いて、

$$p\left(t|x, \mathbf{w}_{\mathrm{ML}}, \beta_{\mathrm{ML}}\right) = \mathcal{N}\left(t|y\left(x, \mathbf{w}_{\mathrm{ML}}\right), \beta_{\mathrm{ML}}^{-1}\right)$$

とかける. ただし、

$$\frac{1}{\beta_{\mathrm{ML}}} = \frac{1}{N} \sum_{n=1}^{N} \left\{y\left(x_n, \mathbf{w}_{\mathrm{ML}}\right) - t_n\right\}^2$$

曲線フィッティングとベイズ推定

ベイズ理論を用いて最尤推定するとどうなるのだろうか.
事前分布を

$$p(\mathbf{w}|\alpha) = \mathcal{N}\left(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbf{I}\right) = \left(\frac{\alpha}{2\pi}\right)^{(M+1)/2} \exp\left\{-\frac{\alpha}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}\right\}$$

とすると、ベイズの定理 $p(\mathbf{w}|\mathbf{x}, \mathbf{t}, \alpha, \beta) \propto p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta)p(\mathbf{w}|\alpha)$ **より、これに対して、対数尤度推定をすると**

$$\frac{\beta}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{\alpha}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}$$

を最小化する問題と等価であることがわかる. これは MAP(maximum posterior) **と呼ばれる.**
上式は、正則化された 2 **乗和誤差関数と等価であることもわかる.**

ベイズ曲線フィッティング

$p(\mathbf{w}|\mathbf{x},\mathbf{t},\alpha,\beta)$ ではなく、$p(t|x,\mathbf{x},\mathbf{t})$ がわかって初めて、予測ができる.α,β を省略して、

$$p(t|x,\mathbf{x},\mathbf{t}) = \int p(t|x,\mathbf{w})p(\mathbf{w}|\mathbf{x},\mathbf{t})\mathrm{d}\mathbf{w}$$

とかける. ガウス分布のすばらしい性質 (興味のある人は 2.3 へ) によって、これらは解析的に計算することができて、
$p(t|x,\mathbf{x},\mathbf{t}) = \mathcal{N}\left(t|m(x),s^2(x)\right)$ とすると、

$$m(x) = \beta\phi(x)^{\mathrm{T}}\mathbf{S}\sum_{n=1}^{N}\phi\left(x_n\right)t_n$$

$$s^2(x) = \beta^{-1} + \phi(x)^{\mathrm{T}}\mathbf{S}\phi(x)$$

ただし、$\phi_i(x) = x^i$ for $i = 0,\ldots,M$ として、

$$\mathbf{S}^{-1} = \alpha\mathbf{I} + \beta\sum_{n=1}^{N}\phi\left(x_n\right)\phi(x)^{\mathrm{T}}$$

ベイズ曲線フィッティング

分散に注目すると赤字の部分がベイズ的な扱いによって、生じた部分.

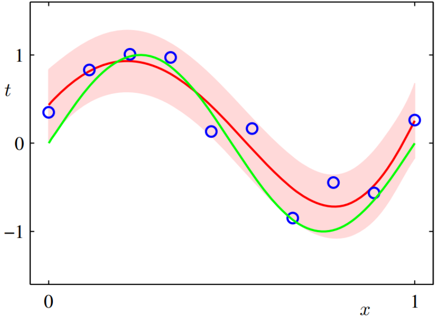

Figure: $M = 9$ **の場合における比較. 赤線：ベイズ、緑：普通の最尤推定**

Contents

1. Gaussian distribution(**ガウス分布**)
2. **ガウス分布と最尤推定**
3. **曲線フィッティング**
4. **ベイズ曲線フィッティング**