Duda & Hart
3.4 LEARNING THE MEAN OF A NORMAL DENSITY

3.4.2 The Univariate Case:$p(x|\mathcal{X})$

3.4.3 The Multivariate Case

作成者: おおつかたく

December 24, 2025

目次

Purpose and Background

目的と背景

- 分散が既知で平均が未知の正規分布に対し、ベイズ推定を用いて平均値を学習する。
- 未知の μ を確率変数とみなし、事前知識（事前分布）とデータ（尤度）を組み合わせて推定する。

3.4.2 1変数の場合: $p(x|\mathcal{X})$

3.4.2 1 変数の場合: $p(x|\mathcal{X})$

事後密度 $p(\mu|\mathcal{X})$ を得たので、残るは「クラス条件付き」密度 $p(x|\mathcal{X})$ を得ることだけである*。式 (14), (15), (19) より、

$$p(x|\mathcal{X}) = \int p(x|\mu)p(\mu|\mathcal{X})\,d\mu$$

* 単純化のためにクラスの区別を省略したが、すべてのサンプルは同一のクラス、例えば ω_1 からのものであり、$p(x|\mathcal{X})$ は実際には $p(x|\omega_1, \mathcal{X}_1)$ である。

3.4.2 1変数の場合: $p(x|\mathcal{X})$ 続き

$$p(x|\mathcal{X}) = \int p(x|\mu)p(\mu|\mathcal{X})\,d\mu$$

$$= \int \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]\frac{1}{\sqrt{2\pi}\sigma_n}\exp\left[-\frac{1}{2}\left(\frac{\mu-\mu_n}{\sigma_n}\right)^2\right]d\mu$$

$$= \frac{1}{2\pi\sigma\sigma_n}\exp\left[-\frac{1}{2}\frac{(x-\mu_n)^2}{\sigma^2+\sigma_n^2}\right]f(\sigma,\sigma_n)$$

ここで、

$$f(\sigma,\sigma_n) = \int \exp\left[-\frac{1}{2}\frac{\sigma^2+\sigma_n^2}{\sigma^2\sigma_n^2}\left(\mu-\frac{\sigma_n^2 x+\sigma^2\mu_n}{\sigma^2+\sigma_n^2}\right)^2\right]d\mu$$

3.4.2 1変数の場合: $p(x|\mathcal{X})$ 続き

つまり、x の関数として見ると、$p(x|\mathcal{X})$ は $\exp\left[-\frac{1}{2}\frac{(x-\mu_n)^2}{\sigma^2+\sigma_n^2}\right]$ に比例し、したがって $p(x|\mathcal{X})$ は平均 μ_n、分散 $\sigma^2+\sigma_n^2$ の正規分布に従う。

$$p(x|\mathcal{X}) \sim N(\mu_n, \sigma^2+\sigma_n^2) \quad (25)$$

3.4.2 1変数の場合: $p(x|\mathcal{X})$ 続き

- 言い換えれば、パラメトリックな形式が $p(x|\mu) \sim N(\mu, \sigma^2)$ であることが既知である「クラス条件付き」密度 $p(x|\mathcal{X})$ を得るには、単に μ を μ_n に、σ^2 を $\sigma^2 + \sigma_n^2$ に置き換えるだけでよい。
- 事実上、条件付き平均 μ_n はあたかも真の平均であるかのように扱われ、既知の分散は、平均 μ に関する正確な知識が欠けていることに起因する x の追加的な不確実性を考慮して増大される。

3.4.2 1 変数の場合: $p(x|\mathcal{X})$ 続き

- これが最終結果である。密度 $p(x|\mathcal{X})$ は求めるクラス条件付き密度 $p(x|\omega_i, \mathcal{X}_i)$ であり、事前確率 $P(\omega_i)$ と合わせることで、ベイズ分類器の設計に必要な確率的情報を与えてくれる。

3.4.3 多変量の場合

3.4.3 多変量の場合

多変量の場合の扱いは、1 変数の場合の直接的な一般化である。したがって、証明の概略のみを簡潔に述べる。以前と同様に、以下を仮定する。

$$p(\vec{x}|\vec{\mu}) \sim N(\vec{\mu}, \mathbf{\Sigma}) \quad (26)$$

$$p(\vec{\mu}) \sim N(\vec{\mu}_0, \mathbf{\Sigma}_0) \quad (27)$$

ここで、$\mathbf{\Sigma}, \mathbf{\Sigma}_0, \vec{\mu}_0$ は既知であると仮定する。

n 個の独立したサンプルの集合 $\mathcal{X} = \{\vec{x}_1, \dots, \vec{x}_n\}$ を観測した後、ベイズの規則を用いて以下を得る。

$$p(\vec{\mu}|\mathcal{X}) = \alpha \prod_{k=1}^{n} p(\vec{x}_k|\vec{\mu})p(\vec{\mu})$$

$$= \alpha' \exp\left[-\frac{1}{2}\left(\vec{\mu}^t(n\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_0^{-1})\vec{\mu} - 2\vec{\mu}^t(\boldsymbol{\Sigma}^{-1}\sum \vec{x}_k + \boldsymbol{\Sigma}_0^{-1}\vec{\mu}_0)\right)\right]$$

これは以下の形式を持つ。

$$p(\vec{\mu}|\mathcal{X}) = \alpha'' \exp\left[-\frac{1}{2}(\vec{\mu} - \vec{\mu}_n)^t\boldsymbol{\Sigma}_n^{-1}(\vec{\mu} - \vec{\mu}_n)\right]$$

3.4.3 多変量の場合 続き

したがって、$p(\vec{\mu}|\mathcal{X}) \sim N(\vec{\mu}_n, \mathbf{\Sigma}_n)$ となり、ここでもまた再生密度（reproducing density）が得られる。係数を等置することで、式 (20) と (21) の類似式を得る。

$$\mathbf{\Sigma}_n^{-1} = n\mathbf{\Sigma}^{-1} + \mathbf{\Sigma}_0^{-1} \quad (28)$$

$$\mathbf{\Sigma}_n^{-1}\vec{\mu}_n = n\mathbf{\Sigma}^{-1}\mathbf{m}_n + \mathbf{\Sigma}_0^{-1}\vec{\mu}_0 \quad (29)$$

ここで $\mathbf{m}_n$ は標本平均である。

$$\mathbf{m}_n = \frac{1}{n}\sum_{k=1}^{n}\vec{x}_k \quad (30)$$

3.4.3 多変量の場合 続き

$\vec{\mu}_n$ と $\mathbf{\Sigma}_n$ に関するこれらの方程式の解は、以下の行列恒等式の知識によって単純化される。

$$(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1} = \mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B} = \mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{A}$$

これは任意の正則な $d \times d$ 行列 $\mathbf{A}$ と $\mathbf{B}$ のペアに対して有効である。少しの操作の後、以下の最終結果を得る。

$$\vec{\mu}_n = \mathbf{\Sigma}_0\left(\mathbf{\Sigma}_0 + \frac{1}{n}\mathbf{\Sigma}\right)^{-1}\mathbf{m}_n + \frac{1}{n}\mathbf{\Sigma}\left(\mathbf{\Sigma}_0 + \frac{1}{n}\mathbf{\Sigma}\right)^{-1}\vec{\mu}_0 \quad (31)$$

$$\mathbf{\Sigma}_n = \mathbf{\Sigma}_0\left(\mathbf{\Sigma}_0 + \frac{1}{n}\mathbf{\Sigma}\right)^{-1}\frac{1}{n}\mathbf{\Sigma} \quad (32)$$

3.4.3 多変量の場合 続き

$p(\vec{x}|\mathcal{X}) \sim N(\vec{\mu}_n, \mathbf{\Sigma} + \mathbf{\Sigma}_n)$ であることの証明は、以前と同様に以下の積分を実行することで得られる。

$$p(\vec{x}|\mathcal{X}) = \int p(\vec{x}|\vec{\mu})p(\vec{\mu}|\mathcal{X})\, d\vec{\mu}$$

3.4.3 多変量の場合 続き

- しかし、この結果は、$\vec{x}$ が 2 つの確率変数の和と見なせることに着目すれば、より少ない労力で得られる。
- すなわち、$p(\vec{\mu}|\mathcal{X}) \sim N(\vec{\mu}_n, \mathbf{\Sigma}_n)$ に従う確率ベクトル $\vec{\mu}$ と、$p(\vec{y}) \sim N(\vec{0}, \mathbf{\Sigma})$ に従う独立した確率ベクトル $\vec{y}$ の和である。
- 独立した正規分布に従う 2 つのベクトルの和は、再び正規分布に従うベクトルとなり、その平均は平均の和、共分散行列は共分散行列の和となるため、

$$p(\vec{x}|\mathcal{X}) \sim N(\vec{\mu}_n, \mathbf{\Sigma} + \mathbf{\Sigma}_n)$$

となり、一般化は完了する。

Conclusion

まとめ

- 今回は、分散が既知で平均が未知の正規分布に対し、ベイズ推定を用いて平均値を学習する方法を説明した。
- 1 変数の場合と多変量の場合の両方で、事後分布とクラス条件付き密度を導出した。