知能システム論第 2 回課題

37186305
航空宇宙工学専攻修士一年
荒居秀尚

2018 年 10 月 11 日

1　宿題 1

A) 次の二乗誤差 $J_1(y)$ を最小化する y を y_1 で表す。

$$y_1 = \arg\min_y J_1(y) \tag{1.1}$$

$$J_1(y) = \int_a^b (x-y)^2 f(x)dx \tag{1.2}$$

この $J_1(y)$ を偏微分したものが 0 になるのが y_1 である。すなわち、

$$\frac{\partial J_1}{\partial y} = 2\int_a^b (x-y)f(x)dx \tag{1.3}$$

$$= 2y\int_a^b f(x)dx - 2\int_a^b xf(x)dx \tag{1.4}$$

これが 0 となる y が y_1 であり、すなわち

$$y_1 = \frac{\int_a^b xf(x)dx}{\int_a^b f(x)dx} \tag{1.5}$$

$f(x)$ が確率密度関数であるため、分母の積分は 1 となり、分子は $E[x]$ であるから、

$$y_1 = E[x] \tag{1.6}$$

である。

B) 次の絶対誤差 $J_2(y)$ を最小にする y を y_2 で表す。

$$y_2 = \arg\min_y J_2(y) \tag{1.7}$$

$$J_2(y) = \int_a^b |x-y|f(x)dx \tag{1.8}$$

このとき、y_2 は X の中央値 (つまり $F(y_2) - 1/2$ であることを示す。

a、b は確率密度分布の境界であることから $a \leq y_2 \leq b$ であるとして考える。

$$J_2(y) = \int_a^b |x-y| f(x)dx \tag{1.9}$$

$$= \int_a^y (y-x)f(x)dx + \int_y^b (x-y)f(x)dx \tag{1.10}$$

$$= y\left[\int_a^y f(x)dx - \int_y^b f(x)dx\right] + \int_y^b xf(x)dx - \int_a^y xf(x)dx \tag{1.11}$$

$$= y\left[2F(y)-1\right] + \int_y^b xf(x)dx - \int_a^y xf(x)dx \tag{1.12}$$

両辺を y で微分して、

$$\frac{\partial J_2(y)}{\partial y} = 2F(y) - 1 + 2yf(y) - 2yf(y) \tag{1.13}$$

$$= 2F(y) - 1 \tag{1.14}$$

よってこれを 0 にする y においては $F(y_2) = \frac{1}{2}$ となるため中央値である。

2　宿題 2

t の 2 次式

$$Q(t) = V(tX+Y) \tag{2.1}$$

$$= E\{tX + Y - E(tX+Y)\}^2 \tag{2.2}$$

$$= t^2V(X) + 2t\mathbf{Cov}(X,Y) + V(Y) \tag{2.3}$$

これは負にならないはずであるから、

$$(\mathbf{Cov}(X,Y))^2 \leq V(X)\cdot V(Y) \tag{2.4}$$

これを解けば、相関係数が

$$-1 \leq \rho_{XY} \leq 1 \tag{2.5}$$

を満たすことがわかる。

3　宿題 3

離散型の場合の証明は同様であるので連続型変数の場合のみ証明を行う。

i.

$$E[XY] = \iint_S xy\cdot f(x,y)dxdy \tag{3.1}$$

$$= \iint_S xy\cdot h(x)g(y)dxdy \tag{3.2}$$

$$= \int xh(x)dx \int yg(y)dy \tag{3.3}$$

$$= E[X]E[Y] \tag{3.4}$$

ここで二行目の変形は独立性の仮定による。

ii.

$$
\begin{align}
M_{X+Y}(t) &= E[e^{t(X+Y)}] \tag{3.5}\\
&= \iint e^{tx}e^{ty}f(x,y)dxdy \tag{3.6}\\
&= \iint e^{tx}e^{ty}h(x)g(y)dxdy \tag{3.7}\\
&= \int e^{tx}h(x)dx \int e^{ty}g(y)dy \tag{3.8}\\
&= M_X(t)M_Y(t) \tag{3.9}
\end{align}
$$

iii. i. と $\mathbf{Cov(X,Y)} = E[XY] - E[X]E[Y]$ より導かれる。

4 宿題 4

Python3.6.6 を用いた。

```python
import numpy as np


def read_data(path):
    with open(path, "r") as f:
        lines = f.readlines()
    lines = [x.replace("\n", "") for x in lines]
    data = np.array([[np.float(x.split(",")[0]),
                      np.float(x.split(",")[1])] for x in lines])
    return data


if __name__ == "__main__":
    base_path = "../../data/correlation"
    data1_path = f"{base_path}1.txt"
    data2_path = f"{base_path}2.txt"
    data3_path = f"{base_path}3.txt"
    data4_path = f"{base_path}4.txt"

    data1 = read_data(data1_path)
    data2 = read_data(data2_path)
    data3 = read_data(data3_path)
    data4 = read_data(data4_path)
    data = [data1, data2, data3, data4]

    for i, d in zip([1, 2, 3, 4], data):
        print(f"Correlation{i}:␣{np.corrcoef(d[:,␣0],␣d[:,␣1])[0,␣1]:.5f}")
```

結果は 4.1 のようになった 相関係数が ほぼ 0 となるようなデータの中にも、データ 3 のように各変数間で何らかの関係が容易には見出せないような場合もあれば、データ 4 のように可視化してみると何らかの関係が見

データ	相関係数
データ 1	0.70302
データ 2	-0.67369
データ 3	0.03691
データ 4	-0.04976

表 4.1: 各データの相関係数

出せそうな場合もあり、相関係数だけを眺めてデータを判断することは危険であるということの好例となっている。