パターン認識と機械学習
1.2.2 期待値と分散

吉田拓夢

期待値
（expectation）

- 確率分布p(x)による関数f(x)への重み付け
  - 離散分布の場合

$$\mathbb{E}[f] = \sum_x p(x) f(x)$$

  - 連続変数の場合

$$\mathbb{E}[f] = \int p(x) f(x) dx$$

期待値の近似

- 確率分布や確率密度から有限個のN点を使ってそれらの有限和で近似できる

$$\mathbb{E}[f] \simeq \frac{1}{N}\sum_{n=1}^{N} f(x_n)$$

- N→∞の極限で近似の厳密化

多変数関数の期待値

- 関数f(x,y)のxの分布に関する平均
- yの関数

$$\mathbb{E}_x[f(x,y)]$$

条件付き期待値（conditional expectation）

- 条件付き分布における条件付き期待値

$$\mathbb{E}_x[f|y] = \sum_x p(x|y) f(x)$$

分散
（variance）

- 定義

$$var[f] = \mathbb{E}\left[(f(x) - \mathbb{E}[f(x)])^2\right]$$

→f(x)の平均値周辺におけるばらつきの尺度

- 2乗の展開

$$var[f] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$$

- x自身の分散

$$var[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2$$

補足

- 分散の定義から2乗の展開(連続変数の場合)

$\mathbb{E}[f(x)]$ は定数となるので、$\mathbb{E}[f(x)] = c$ とおく.

$$
\begin{aligned}
var[f] &= \mathbb{E}\left[(f(x) - \mathbb{E}[f(x)])^2\right] \\
&= \mathbb{E}\left[(f(x) - c)^2\right] \\
&= \int p(x)\{f(x)^2 - 2cf(x) + c^2\}dx \\
&= \int p(x)f(x)^2 dx - 2c \int p(x)f(x)dx + c^2 \int p(x)dx \\
&= \mathbb{E}[f(x)^2] - 2c\mathbb{E}[f(x)] + c^2 \cdot 1 \\
&= \mathbb{E}[f(x)^2] - 2c^2 + c^2 \\
&= \mathbb{E}[f(x)^2] - c^2 \\
&= \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2
\end{aligned}
$$

共分散
（covariance）

- 定義

$$\begin{aligned} cov[x,y] &= \mathbb{E}_{x,y}\left[\{x-\mathbb{E}[x]\}\{y-\mathbb{E}[y]\}\right] \\ &= \mathbb{E}_{x,y}[xy]-\mathbb{E}[x]\mathbb{E}[y] \end{aligned}$$

2つの確率変数x,yが同時に変動する度合い
xとyが独立→共分散は0

ベクトルの共分散

- 定義

$$
\begin{aligned}
cov[\mathbf{x}, \mathbf{y}] &= \mathbb{E}_{\mathbf{x},\mathbf{y}}\left[\{\mathbf{x} - \mathbb{E}[\mathbf{x}]\}\{\mathbf{y}^{\mathrm{T}} - \mathbb{E}[\mathbf{y}^{\mathrm{T}}]\}\right] \\
&= \mathbb{E}_{\mathbf{x},\mathbf{y}}[\mathbf{x}\mathbf{y}^{\mathrm{T}}] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}^{\mathrm{T}}]
\end{aligned}
$$

- ベクトルXの成分間の共分散

$$cov[\mathbf{x}] \equiv cov[\mathbf{x}, \mathbf{x}]$$