パターン認識と機械学習

２．２　多値変数

松本良太

多値変数

- 多値変数について
- 1対K法

多値変数について

二値変数は、2つのとりうる値のうち1つをとる量を記述するのに利用できる。

↓

実際には、もっと多くの状態のうちから1つを取るような離散変数を扱う必要がある。（多値変数）

1対K法

変数は、要素のひとつx_kが1で、残りすべては0であるような、

K次元ベクトルxで表される。

例：K=6種類の状態を取りうる変数で、その中のx_3が1となった観測値xは、

x＝ $(0,0,1,0,0,0)^T$

なお、このようなベクトルは$\sum_{k=1}^{K} x_k = 1$ を満たす。

1対K法

x_k ＝1となる確率をパラメータμ_kで表すと、xの分布は

$$p(x \mid \mu) = \prod_{k=1}^{K} \mu_k^{x^k}$$

ただし、$\mu = (\mu_1, \ldots \mu_K)^T$

$$\mu_k \geq 0$$

$$\sum_k \mu_k = 1$$

1対K法

N個の独立な観測値x1・・・xNのデータ集合Dを考える

このときの尤度関数は、

$$p(D|\mu) = \prod_{n=1}^{N}\prod_{k=1}^{K}\mu_k^{x_{nk}} = \prod_{k=1}^{K}\mu_k^{(\sum_n x_{nk})} = \prod_{k=1}^{K}\mu_k^{m_k}$$

$$m_k = \sum_n x_{nk}$$

これは、$x_k = 1$となる観測値の数を表し、

この分布の十分統計量と呼ばれる。

１対K法

・μの最尤推定解を求める。
μkについてlnp(D|μ)を最大化する。
ラグランジュ乗数λを用いて下の式を最大化する。

$$\sum_{k=1}^{K} m_k \ln \mu_k + \lambda\left(\sum_{k=1}^{K} \mu_k - 1\right)$$

上の式について、μkについての導関数を0とおくと、
$\mu_k = -m_k / \lambda$
を得る。

１対K法

・続き

この分布には $\sum_n \mu_k = 1$ という制約があるので、

これに先ほどの値を代入する。

すると、λ ＝－Nと解けるので、

最尤推定解は、 $\mu_k^{ML} = \frac{m_k}{N}$

これは、N個の観測値のうち χ k＝１であるものの割合になっている。

１対K法

パラメータμと観測値の総数Nが与えられた条件の元で、ｍ1・・・mKの同時確率を考える。

$$Mult(m_1....m_K \mid \mu, N) = \binom{N}{m_1 m_2 ... m_K} \prod_{k=1}^{K} \mu_k^{m_k}$$

これは、多項分布として知られるものである。
正規化係数は、

$$\binom{N}{m_1 m_2 ... m_k} = \frac{N!}{m_1! m_2! .... m_k!}$$

で与えられる。ただし、

$$\sum_{k=1}^{K} m_k = N$$