

パターン認識と機械学習

2.2 多値変数

吉田拓夢

1対K法

- 要素の1つ x_k が1で他はすべて0のK次元ベクトル \mathbf{x}

例:K=6、 $x_3=1$ の状態の観測値 \mathbf{x}

$$\mathbf{x} = (0, 0, 1, 0, 0, 0)^T$$

- 以下を満たす

$$\sum_{k=1}^K x_k = 1$$

確率分布

- $x_k=1$ となる確率をパラメータ μ_k とした x の分布

$$p(\mathbf{x}|\boldsymbol{\mu}) = \prod_{k=1}^K \mu_k^{x_k} \quad \boldsymbol{\mu} = (\mu_1, \dots, \mu_K)^T$$

- μ_k は確率を表すので以下を満たす

$$\sum_k \mu_k = 1$$

$$\mu_k \geq 0$$

確率分布

$$p(\mathbf{x}|\boldsymbol{\mu}) = \prod_{k=1}^K \mu_k^{x_k}$$

←この分布はベルヌーイ分布の一般化
(2種類以上の出力)とみなせる

$$\sum_{\mathbf{x}} p(\mathbf{x}|\boldsymbol{\mu}) = \sum_{k=1}^K \mu_k = 1$$

$$\mathbb{E}[\mathbf{x}|\boldsymbol{\mu}] = \sum_{\mathbf{x}} p(\mathbf{x}|\boldsymbol{\mu})\mathbf{x} = (\mu_1, \dots, \mu_K)^T = \boldsymbol{\mu}$$

尤度関数

- N 個の独立な観測値 x_1, \dots, x_N のデータ集合 D を考えた時の尤度関数は

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

- 以下は $x_{nk}=1$ となる観測値の数を表す
→十分統計量(sufficient statistic)

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

最尤推定解

- μ の最尤推定解は μ_k の総和が1になることを考慮して μ_k について $\ln p(\mathcal{D}|\mu)$ を最大化

→ラグランジュ乗数 λ を用いて以下を最大化

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

→上記の μ_k についての導関数を0とすると

$$\mu_k = -m_k / \lambda$$

最尤推定解

$\mu_k = -m_k / \lambda$ を制約 $\sum_k \mu_k = 1$ に代入

→ラグランジュ乗数 $\lambda=-N$ と解ける.よって、最尤推定解は

$$\mu^{\text{ML}} = \frac{m_k}{N}$$

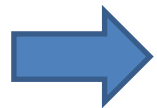


N個の観測値のうち $x_k=1$ であるものの割合

同時確率

- パラメータ μ と観測値の総数 N が与えられた条件下で m_1, \dots, m_K の同時確率は

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



多項分布(multinomial distribution)

- 正規化係数は、 N 個の対象を大きさが m_1, \dots, m_K の K 個のグループに分割する場合の数に相当

$$\binom{N}{m_1, m_2, \dots, m_K} = \frac{N!}{m_1! m_2! \dots m_K!} \quad \sum_{k=1}^K m_k = N$$