

scikit-learn ゼミ

1.7. Naive Bayes

1.7.1. Gaussian Naive Bayes

小野寺喜行

Naive Bayes

- ・教師あり学習アルゴリズム
- ・特徴それぞれが独立という単純な仮定とベイズの定理を用いる
- ・多くのタスクでとてもよく機能する
- ・少しのトレーニングデータでパラメータ推定できる
- ・より高度な手法と比較して高速

Naive Bayes

ベイズの定理は以下の関係を示す

$$P(y | x_1, \dots, x_n) = \frac{P(y)P(x_1, \dots, x_n | y)}{P(x_1, \dots, x_n)}$$

(クラス変数: y , 特徴ベクトル: x_1, \dots, x_n)

Naive Bayes

単純な独立仮定によって

$$P(x_i | y, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = P(x_i | y)$$

次の様に単純化

$$P(y | x_1, \dots, x_n) = \frac{P(y) \prod_{i=1}^n P(x_i | y)}{P(x_1, \dots, x_n)}$$

Naive Bayes

$P(x_1, \dots, x_n)$ が恒常的に入力を与えられるとき

$$P(y | x_1, \dots, x_n) \propto P(y) \prod_{i=1}^n P(x_i | y)$$



$$\hat{y} = \arg \max_y P(y) \prod_{i=1}^n P(x_i | y)$$

* $P(y)$ と $P(x_i|y)$ の推定には、
Maximum A Posteriori (MAP) 評価が使える

Gaussian Naive Bayes

GaussianNB : Gaussian Naive Bayesアルゴリズムを実装

$$P(x_i | y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left\{-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right\}$$

(パラメータ σ_y と μ_y は最尤法で推定)

```
>>> from sklearn import datasets
>>> iris = datasets.load_iris()
>>> from sklearn.naive_bayes import GaussianNB
>>> gnb = GaussianNB()
>>> y_pred = gnb.fit(iris.data, iris.target).predict(iris.data)
>>> print("Number of mislabeled points out of a total %d points : %d"
...      % (iris.data.shape[0], (iris.target != y_pred).sum()))
Number of mislabeled points out of a total 150 points : 6
```