

単純ベイズ分類器
ナイーブベイズ
Naïve Bayes Classifier

明治大学 理工学部 応用化学科
データ化学工学研究室 金子 弘昌

ナイーブベイズとは？

- ✓ クラス分類手法の一つであり、多クラス分類も可能
- ✓ ベイズの定理を利用
- ✓ 生成モデルの一つであり、クラス分類の結果が確率として得られる
- ✓ それぞれの説明変数 (入力変数・記述子・特徴量) は独立していること (正確にいうと条件付き独立性) を仮定
- ✓ 説明変数の分布 (正確には、クラスが与えられたときの説明変数の分布) を仮定する必要がある
 - 一般的には、正規分布かベルヌーイ分布

ナイーブベイズで求めたいもの

$$p(y | X)$$

- ✓ X が与えられたときの、 y の確率分布
 - X : 説明変数 (入力変数・記述子・特徴量)
 - y : 目的変数 (クラス)

- ✓ X を入力すると、各クラスの確率が出てくる

ベイズの定理

$$p(y | X) = \frac{p(y) p(X | y)}{p(X)}$$

✓ $p(y)$: y の事前確率分布

- たとえば、100 個のサンプルがあって、そのうち 20 個がクラスA, 30 個がクラスB, 50 個がクラスC のとき、

- $p(y = A) = 20/100 = 0.2$

- $p(y = B) = 30/100 = 0.3$

- $p(y = C) = 50/100 = 0.5$

✓ $p(X | y)$: y が与えられたときの、 X の確率分布

✓ $p(X)$: X の事前確率分布

大事なものは分子だけ

$$p(y|X) = \frac{p(y)p(X|y)}{p(X)} \quad \Rightarrow \quad p(y|X) \propto p(y)p(X|y)$$

✓ $p(y|X)$ は $p(y)p(X|y)$ に比例

✓ たとえばクラス A, B, C があるとき、

- $p(y=A)p(X|y=A)$
- $p(y=B)p(X|y=B)$
- $p(y=C)p(X|y=C)$

を求めてから、確率の和が 1 になるように、それぞれ
 $p(y=A)p(X|y=A) + p(y=B)p(X|y=B) +$
 $p(y=C)p(X|y=C)$ で割ればよい

説明変数の間の独立性

✓説明変数の数を m 個とする

$$X = [x_1 \quad x_2 \quad \cdots \quad x_m]$$

✓ y のクラスが与えられたとき、説明変数はそれぞれ独立していると仮定

✓つまり、

$$\begin{aligned} p(X | y) &= p(x_1 | y) \times p(x_2 | y) \times \cdots \times p(x_m | y) \\ &= \prod_{i=1}^m p(x_i | y) \end{aligned}$$

✓よって、 $p(y | X) \propto p(y) p(X | y) = p(y) \prod_{i=1}^m p(x_i | y)$

確率分布を仮定

✓ $p(x_i | y)$ をどういう確率分布にするか？

- 正規分布 (ガウス分布)
 - よく用いられる

$$p(x_i | y) = \frac{1}{\sqrt{2\pi\sigma_{x_i,y}^2}} \exp \left\{ -\frac{(x_i - \mu_{x_i,y})^2}{2\sigma_{x_i,y}^2} \right\}$$

- ベルヌーイ分布
 - x_i が 0 もしくは 1 の変数のとき

$$p(x_i | y) = q_y^{x_i} (1 - q_y)^{1-x_i}$$

確率分布のパラメータをどうするか？

✓正規分布 (ガウス分布)

- y のクラスごとに、説明変数 x_i ごとに、データセットから平均 $\mu_{xi,y}$, 標準偏差 $\sigma_{xi,y}$ を計算

✓ベルヌーイ分布

- y のクラスごとに、説明変数 x_i ごとに、データセットから $x_i = 1$ のときにクラスに属するサンプル数を全サンプル数で割って q_y とする

どのようにパラメータが導出されるか？

✓ 尤度関数 L を

$$L = \prod_{j=1}^n p\left(y^{(j)}\right) \prod_{i=1}^m p\left(x_i^{(j)} \mid y^{(j)}\right)$$

としたときに、最尤推定法により確率分布のパラメータが導出される

- n : サンプル数

✓ 最尤推定法についてはこちら

- <https://datachemeng.com/maximumlikelihoodestimation/>