知能機械と自然言語処理
知能機械部　第9回

ソフトウェア情報学部

Goutam Chakraborty

ナイーブベイズ (Naive Bayes’)

ナイーブベイズ (Naive Bayes’)は下記の問題に有効である:
- **特色の数が多い**
- **特色どうしの相関がない/少ない**
- **サンプルの数が少ない**

- 多次元（特色の数が多い場合）相関行列とそれの逆行列の計算は大変なので、同じくクラス条件付き確率 $P(\vec{X}\mid C_i)$ を求める事も大変です。
- しかし、特色どうしがお互い独立している場合にナイーブベイズが使えます。
- 特色の数が増えれば増えるほど正しく分類するために膨大のデータが必要です。例えば1次元の分類にN個のデータが必要な時、d 次元だと N^d 個のデータが必要です。しかし、ナイーブベイスを使うと各特色ごとに確率を求めるので、データが少なくても分類が可能です。
- 特にText Mining等の問題では、次元数が多い（100以上）ので、ナイーブベイスはよく使われます。

特色独立された場合の相関行率

$$\begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 & 0 \\ 0 & a_{22} & 0 & & & 0 \\ 0 & \ddots & a_{33} & 0 & & \vdots \\ \vdots & & 0 & a_{44} & \ddots & 0 \\ 0 & & \cdots & 0 & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 0 & a_{dd} \end{bmatrix}$$

全ての特色が独立されている場合、相関行列の対角線以外の成分は0です。
その為 $P(\vec{X}\mid C_i)$ を求めるときに相関行列を求める必要がないです。

ナイーブベイズを用いた条件付き確率の求め方

d次元の正規分布のデータの確率密度関数

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})\right) \quad [1]$$

よって $P(\vec{X}\mid C_i)$ を求めるために $\Sigma, |\Sigma|, \Sigma^{-1}$ を求めなければなりません。

しかし、ナイーブベイズを使用することで $P(\vec{X}\mid C_i)$ は下記の式で求められます。

$$P(\vec{X}\mid C_i) = \prod_{j=1}^{d} P(x_j\mid C_i) = P(x_1\mid C_i)\times P(x_2\mid C_i)\times\cdots\times P(x_d\mid C_i) \quad [2]$$

d は特色の数です

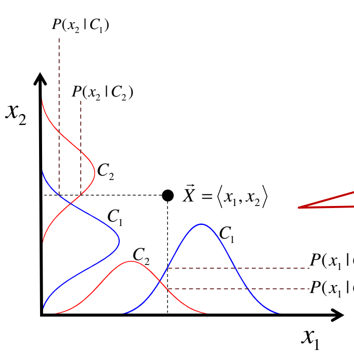
計算量は[1]より[2]の方が遥かに軽いです。

Class数は *k* です。*i* が1 から *k* までの $P(\vec{X}\mid C_1)$、$P(\vec{X}\mid C_2)$、$P(\vec{X}\mid C_3)$ ….. $P(\vec{X}\mid C_k)$ *を先ず求めて、式[2]を用いて* $P(\vec{X}\mid C_i)$ *を求める事です。最後に事後確率は*

$$P(C_i\mid \vec{X}) = P(\vec{X}\mid C_i)\times P(C_i) = [P(x_1\mid C_i)\times P(x_2\mid C_i)\times\cdots\times P(x_d\mid C_i)]\times P(C_i) \quad [3]$$

Class $i = 1 \sim k$ について、上の式[3]でそれぞれの確率を求めて、最大となるクラスを求める。

ナイーブベイズを用いた条件付き確率の求め方



2次元の場合

$$p(\vec{x}\mid C_1) = p(x_1\mid C_1)\times p(x_2\mid C_1)$$
$$p(\vec{x}\mid C_2) = p(x_1\mid C_2)\times p(x_2\mid C_2)$$

$$P(C_1\mid \vec{X}) = P(\vec{X}\mid C_1)\times P(C_1)$$
$$P(C_2\mid \vec{X}) = P(\vec{X}\mid C_2)\times P(C_2)$$

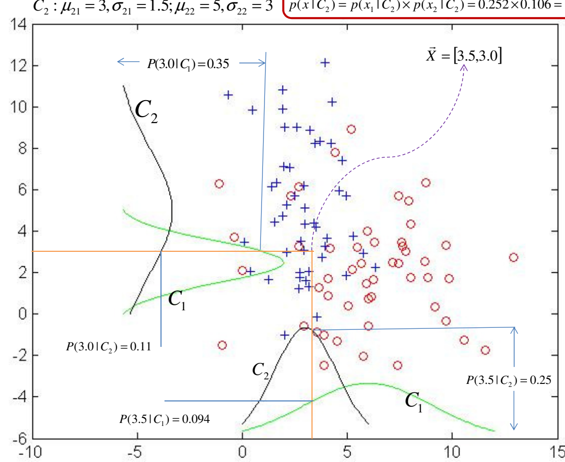
値が大きい方のクラスになる。

独立（相関が無い）のとき、
Naive-Bayes と Bayes’の結果は同じになる

Naïve Bayes’

$C_1 : \mu_{11} = 6, \sigma_{11} = 3; \mu_{12} = 2.5, \sigma_{12} = 1$ $\quad p(\vec{x}\mid C_1) = p(x_1\mid C_1)\times p(x_2\mid C_1) = 0.094\times 0.353 = 0.0332$

$C_2 : \mu_{21} = 3, \sigma_{21} = 1.5; \mu_{22} = 5, \sigma_{22} = 3$ $\quad p(\vec{x}\mid C_2) = p(x_1\mid C_2)\times p(x_2\mid C_2) = 0.252\times 0.106 = 0.0267$



An Application Example