ナイーブなベイズ(補足)

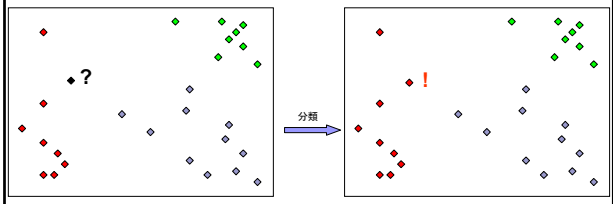
慶應義塾大学理工学部
櫻井 彰人

はじめに

- 以下のスライドは、ナイーブベイズに対する補足説明資料です。
- 説明が不要な方は、次の事項を調べて、自習してください。
- モンティ・ホール問題(友野典男「行動経済学」にも出ている)を、「知っている」「情報を持っている」等の主観的な表現を交えずに、純粋に頻度だけに基づいて、式を用いて回答するとともに、その妥当性を説明してください。
  - 参考:
    - 日本語版Wikipedia の「ベイズ推定」中の「モンティ・ホール問題」
    - 日本語版Wikipedia の「モンティ・ホール問題」
  - ある行動の結果「確率が変わる」という表現も使わないで下さい。確率は、情報が増えた(?)からといって、変わりません。(予測正解頻度を最大にするという目的のもと)考えるべき条件付確率が、異なるものになるだけです。すなわち、あまたある条件付確率のなかで、考えるべきものが変わるだけです。

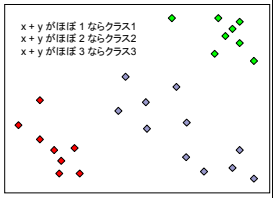
分類とは

- 分類: 仲間集めは、予めされている。課題は、新しいデータが来たときに、どのグループ(クラスという)に入れるかを決めるのだが、その正確さが要求される
  - 例:病気の診断。貸し倒れそうか否かの判断
    - 病気の診断: 過去の症例がある(知識。データ)。患者と思しき人が来る。症状を見る。それをもとに、病気か否か、病気ならその原因は何かを診断する。
    - 貸し倒れ: 住宅ローンを組に来た。収入を見る。職業を見る。家族構成を見る。過去のクレジット返済履歴を見る。物件の担保価値を見る。さあ、貸すか貸さぬか。貸すならいくらまで貸せるか。



分類の方法

- 実にたくさんの方法がある
- 回帰だって使える
- 決定木だって当然



- 今回は、文書分類にも使える naïve Bayes 法を考える
  - 確率的に考える
  - 現実のデータは不確実である。そしてその原因は、確率的現象にある、と考える。

Naïve Bayes 分類器の特徴

- 単純だが(だから?)よく知られた分類方法
  - 単純な割には高精度
  - 単純なだけに、高速
- Bayes 定理 + 仮定 *条件付独立*
  - 実際には成り立たないことが多い仮定
  - それにも関わらず, 実際にはしばしばうまくいく
- 成功事例:
  - 文書分類
  - 診断

Bayes 分類（最尤推定）

- 前提: 原因（クラス）が $c_1, c_2, \ldots, c_k$ であるとし、このいずれかから、サンプルが x が生成されたとする。
- 課題: 原因を推定する方法を考えよ。
- 解: 仮に各原因（クラス）がサンプル x を生成する確率 $\mathrm{Prob}(x \mid c_i)$ が分かっているとする。
  - 仮の課題と解: 事象 $e_1, e_2, \ldots, e_k$ のうちの唯一つが発生した（どれが発生したかは不明）という仮定のもと、発生事象を推定せよという課題に対しては、確率最大の事象が発生したと考えるのが妥当。
- 従って、$\mathrm{argmax}_i \mathrm{Prob}(x \mid c_i)$ とするのが妥当

Bayes 分類（最大事後確率推定）

- 前提: 原因（クラス）が $c_1, c_2, \ldots, c_k$ であるとし、このいずれかから、サンプルが x が生成されたとする。
- 課題: 原因を推定する方法を考えよ。
- 解: 仮に各原因（クラス）とサンプル x が発生する同時確率 $\mathrm{Prob}(x, c_i)$ が分かっているとする（実際は x が定数になっているので、確率変数は分類クラスを値とするもののみ）。
  - 仮の課題と解: 事象 $e_1, e_2, \ldots, e_k$ のうちの唯一つが発生した（どれが発生したかは不明）という仮定のもと、発生事象を推定せよという課題に対しては、確率最大の事象が発生したと考えるのが妥当。
- 従って、$\mathrm{argmax}_i \mathrm{Prob}(x, c_i)$ とするのが妥当

最尤推定 or 最大事後確率推定？

- その前に、違いは？
  - 最尤推定: $\mathrm{argmax}_i \mathrm{Prob}(x \mid c_i)$
  - 事後確率最大: $\mathrm{argmax}_i \mathrm{Prob}(x, c_i)$
- $\mathrm{Prob}(x, c_i) = \mathrm{Prob}(x \mid c_i) \mathrm{Prob}(c_i)$ であるから、
  - 最尤推定: $\mathrm{argmax}_i \mathrm{Prob}(x \mid c_i)$
  - 事後確率最大: $\mathrm{argmax}_i \mathrm{Prob}(x \mid c_i) \mathrm{Prob}(c_i)$

ところで、「条件付確率」は覚えていますか？

用語: 条件付確率

- E.g., P(*cavity* | *toothache*) = 0.8
  i.e., (“歯痛がある”ことを *toothache* で表すと) *toothache* が知っていることのすべてであるとき、cavity となる確率, または cavity であると信ずる信頼度・信念
- これは、次の規則が成立する確率・信頼度・信念と考えることもできる
  if *Toothache=true* then *Cavity=true*
- **条件付確率の定義**:
  $P(a \mid b) = P(a \wedge b) / P(b)$ if $P(b) > 0$
- 確率の積法則 上記定義の（単なる）言い換えである:
  $P(a \wedge b) = P(a \mid b) P(b) = P(b \mid a) P(a)$

ただし、Bayes的アプローチでは、条件付確率を基本におくと考えてよい

用語: ベイズ規則 Bayes' Rule

- 確率の積の法則
  $P(a \wedge b) = P(a \mid b) P(b) = P(b \mid a) P(a)$
  Bayes' rule:
  $P(a \mid b) = P(b \mid a) P(a) / P(b)$
- 因果関係の確率から診断確率をえるのに有用:
  - $P(\mathrm{Cause} \mid \mathrm{Effect}) = P(\mathrm{Effect} \mid \mathrm{Cause}) P(\mathrm{Cause}) / P(\mathrm{Effect})$
- ベイズ規則は実応用上有用である（非常によく使う）. ベイズ規則の右辺3項の予測が結構まともにでき、しかも、左辺の項を知りたいことが多いからである.

最尤推定 or 最大事後確率推定？

- その前に、違いは？
  - 最尤推定: $\mathrm{argmax}_i \mathrm{Prob}(x \mid c_i)$
  - 事後確率最大: $\mathrm{argmax}_i \mathrm{Prob}(x, c_i)$
- $\mathrm{Prob}(x, c_i) = \mathrm{Prob}(x \mid c_i) \mathrm{Prob}(c_i)$ であるから、
  - 最尤推定: $\mathrm{argmax}_i \mathrm{Prob}(x \mid c_i)$
  - 事後確率最大: $\mathrm{argmax}_i \mathrm{Prob}(x \mid c_i) \mathrm{Prob}(c_i)$
- つまり違いは、$\mathrm{Prob}(c_i)$ があるかないか。
  - 仮に$\mathrm{Prob}(c_i)$ の違いが小さければ、（思想は別だが）結果は同じとなる可能性が高い

例: ベイズ規則を使う

例えば,

- ある医者が、髄膜炎が原因で 50% の割合で髄膜炎患者は肩こりを訴えることを知っていたとする.
- その医者は, さらに, 次のことを知っているとする:
  ある患者が髄膜炎である事前確率は 1/50,000 であり
  任意の患者が肩こりを訴える事前確率は 1/20 である.
- 次のように記号を定める
  Meningitis="髄膜炎あり", StiffNeck="肩こりあり"

ベイズ規則 (続き)

P(StiffNeck=true | Meningitis=true) = 0.5
P(Meningitis=true) = 1/50000
P(StiffNeck=true) = 1/20

P(Meningitis=true | StiffNeck=true)
= P(StiffNeck=true | Meningitis=true) P(Meningitis=true) / P(StiffNeck=true)
= (0.5) x (1/50000) / (1/20)
= 0.0002

すなわち, 肩こりのある患者のたった 1 / 5000 が髄膜炎である.
また,
P(Meningitis=false | StiffNeck=true)
= P(StiffNeck=false | Meningitis=false) P(Meningitis=false) / P(StiffNeck=true)
となるゆえ 1/ P(StiffNeck=true) は共通項である.

ベイズ規則 (続き)

- 勿論, 医者は, 肩こりがあった場合 1 / 5000 の率で髄膜炎の可能性ありと知っているということはできよう;
  - つまり, 医者は症状(結果)から原因を得る診断行為に関する量的情報を持っているということもできよう.
  - そのような医者には勿論, ベイズ規則は不要である?!
- しかし残念なことに, 診断知識は因果知識に比べて、ずっと脆弱である.
- 例えば, 髄膜炎の流行が突然始まったとしよう. 事前知識, P(Meningitis=true), は上昇する.
  - 診断確率P(Meningitis=true|StiffNeck=true) を, 流行前の統計的観測から導出していた医者は, この値を(流行に合せ)更新する方法を持ち合わせないことになる.
  - 診断確率を必要となる度に他の3値から計算している医者は P(Meningitis=true|StiffNeck=true) は P(Meningitis=true) に比例して上昇させることができる.
- 言わずもがなであるが, P(StiffNeck=true|Meningitis=true) は流行には影響されない. 単に髄膜炎がどう影響を及ぼすかを示しているだけであるから.

変数が増えても大丈夫?

$$P(cause|effect_1,effect_2)=\frac{P(cause,effect_1,effect_2)}{P(effect_1,effect_2)}=\alpha\, P(cause,effect_1,effect_2)$$

$$=\alpha\, P(effect_1,effect_2,cause)$$
$$=\alpha\, P(effect_1|effect_2,cause)P(effect_2,cause)$$
$$=\alpha\, P(effect_1|effect_2,cause)P(effect_2|cause)P(cause)$$

- $effect_1$ が $effect_2$ に対して独立でなくとも, $cause$ が与えられれば独立となる可能性がある.
- 例えば, 次を考えてみよう
  - $effect_1$ が "読解力"
  - $effect_2$ が "腕の長さ"
- 確かに "読解力" は "腕の長さ" に依存している. だって, (子供を考えてごらん)腕が長ければ腕の短い人より年上なんだから, 読解力は上だよね....
- しかし, 無論, $cause$ のところに '年齢' が与えられれば, "読解力" は "腕の長さ" に依存はしない.

Naive Bayes – 条件付独立性を仮定

$$P(cause|effect_1,effect_2)$$
$$=\alpha\, P(effect_1|effect_2,cause)P(effect_2|cause)P(cause)$$
$$=\alpha\, P(effect_1|cause)P(effect_2|cause)P(cause)$$

$$P(cause|effect_1,....,effect_n)=\alpha\, P(effect_1|cause)...P(effect_n|cause)P(cause)$$

- 二つの仮定を考えよう:
  - 属性はどれも, 区別なく同様に重要である
  - クラス値が与えられれば, 条件付独立である
- この仮定のもとでは, ある属性の値がわかっても, 他の属性に関する知識は何も得られないことになる(クラスがわかっているとき)
- 実世界でこの仮定が成立するなんてとても考えられない. が, 実応用では, この枠組みが結構うまく働くのである!

Naïve Bayes を分類に用いる

- 分類規則の学習: 新事例が与えられたとき, クラスは?
  - 証拠 E = **事例**
  - 原因 H = 当の事例に対する**クラス**値
- Naïve Bayes 仮定: 証拠は(統計的に)独立な部分に分解できる (i.e. 事例の属性!)

P(H | E) = P(E | H) P(H) / P(E)
= P(E_1, E_2,... ,E_n|H) P(H) / P(E)
= P(E_1|H) P(E_2|H)... P(E_n|H) P(H) / P(E)