Classification: Naive Bayes Classifiers

Ninghao Liu

University of Georgia

March 14, 2023

Naive Bayes Classifiers: An Example

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

How to predict the label given a new instance?

$$\boldsymbol{x} : [Refund = No, Married, Income = 120K] \tag{1}$$

Naive Bayes Classifiers

Approach the classification problem from the probabilistic perspective.

How to formally define the problem?

- Given an instance $\boldsymbol{x} = [x_1, x_2, ..., x_D]$, choose y that maximizes

$$P(y|x_1, x_2, ..., x_D).$$

- A model based on *posterior probability*.

Can we estimate $P(y|x_1, x_2, ..., x_D)$ directly from data?

Naive Bayes Classifiers: Bayes theorem

- Approach:
  - Compute $P(y|x_1, x_2, ..., x_D)$ using the **Bayes theorem**

$$P(y|x_1, x_2, ..., x_D) = \frac{P(x_1, x_2, ..., x_D|y)P(y)}{P(x_1, x_2, ..., x_D)}$$

    for each choice of $y \in \{1, 2, ..., C\}$.
  - Choose the y that maximizes $P(y|x_1, x_2, ..., x_D)$.
  - Equivalent to choosing value of y that maximizes

$$P(x_1, x_2, ..., x_D|y)P(y) \text{ or } P(x_1, x_2, ..., x_D, y)$$

- How to compute $P(x_1, x_2, ..., x_D|y)$?

Naive Bayes Classifiers: Bayes theorem

Question: Can we directly compute $P(x_1, x_2, ..., x_D | y)$?

Naive Bayes Classifiers

Compute $P(x_1, x_2, ..., x_D|y)$ for an arbitrary instance $\boldsymbol{x}$.

- Think about a binary classification scenario, where $D = 3$ and each feature x_j has 2 possible values.
    - What type of intermediate data to store?
    - How large is the intermediate data?

Naive Bayes Classifiers: Challenges

Challenges for computing $P(x_1, x_2, ..., x_D|y)$.

- Think about a binary classification scenario, where $D = 3$ and each feature x_j has 2 possible values.
  - What type of intermediate data to store?
  - How large is the intermediate data?
- What about when $D = 100$ and each feature x_j has 16 possible values?
  - What is the challenge?

Naive Bayes Classifiers: Conditional Independence

Definition: **Conditional Independence**.

Features in $\boldsymbol{x}$ are mutually independent, conditional on y. That is,

$$P(x_1, x_2, ..., x_D|y) = P(x_1|y) \cdot P(x_2|y) \cdot \cdots \cdot P(x_D|y).$$

Naive Bayes Classifiers: Conditional Independence

Definition: **Conditional Independence**.

Features in $\boldsymbol{x}$ are mutually independent, conditional on y. That is,

$$P(x_1, x_2, \ldots, x_D | y) = P(x_1|y) \cdot P(x_2|y) \cdot \cdots \cdot P(x_D|y).$$

- Example: Arm length and reading skills
  - Young child has shorter arm length and limited reading skills, compared to adults
  - If age is fixed, no apparent relationship between arm length and reading skills
  - Arm length and reading skills are conditionally independent given age

Naive Bayes Classifiers: Framework

1. Compute $P(x_j|y)$ and $P(y)$ for all cases given the training data.
2. Assume conditional independence:

$$P(x_1, x_2, ..., x_D|y) = P(x_1|y) \cdot P(x_2|y) \cdot \dots \cdot P(x_D|y).$$

3. A new instance $\boldsymbol{x}$ is classified to y_c if

$$P(y_c) \cdot \prod_{j=1}^{D} P(x_j|y_c)$$

is maximal.

Naive Bayes Classifiers: Computation Cost

Compute $P(x_1|y) \cdot P(x_2|y) \cdot \cdots \cdot P(x_D|y)$.

- Think about a binary classification scenario, where $D = 3$ and each feature x_j has 2 possible values.
    - What type of intermediate data to store?
    - How large is the intermediate data?
- What about when $D = 100$ and each feature x_j has 16 possible values?

Naive Bayes Classifiers: An Example

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- Class: $P(Y) = N_c/N$
  - e.g., $P(No) = 7/10$,
    $P(Yes) = 3/10$

Let's work out all $P(X|Y)$

- For discrete attributes:

$$P(X_i \mid Y_k) = |X_{ik}| / N_c$$

  - where $|X_{ik}|$ is number of instances having attribute value X_i and belonging to class Y_k
  - Examples:

    P(Status=Married|No) = 4/7
    P(Refund=Yes|Yes)=0

Naive Bayes Classifiers: An Example

- For continuous attributes:
  - Discretization: Partition the range into bins:
    - Replace continuous value with bin value
  - Probability density estimation:
    - Assume an attribute follows a normal distribution
    - Use data to estimate parameters of distribution (e.g., mean and standard deviation)
    - Once probability distribution is known, use it to estimate the conditional probability $P(X_i|Y)$

Naive Bayes Classifiers: An Example

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- Normal distribution:

$$P(X_i \mid Y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{(X_i-\mu_{ij})^2}{2\sigma_{ij}^2}}$$

  - One for each (X_i, Y_i) pair
- For (Income, Class=No):
  - If Class=No
    - sample mean = 110
    - sample variance = 2975

$$P(Income = 120 \mid No) = \frac{1}{\sqrt{2\pi}(54.54)} e^{-\frac{(120-110)^2}{2(2975)}} = 0.0072$$

Naive Bayes Classifiers: An Example

Given a Test Record:

$$X = (\text{Refund} = \text{No}, \text{Married}, \text{Income} = 120\text{K})$$

naive Bayes Classifier:

P(Refund=Yes|No) = 3/7
P(Refund=No|No) = 4/7
P(Refund=Yes|Yes) = 0
P(Refund=No|Yes) = 1
P(Marital Status=Single|No) = 2/7
P(Marital Status=Divorced|No)=1/7
P(Marital Status=Married|No) = 4/7
P(Marital Status=Single|Yes) = 2/3
P(Marital Status=Divorced|Yes)=1/3
P(Marital Status=Married|Yes) = 0

For taxable income:
If class=No: sample mean=110
sample variance=2975
If class=Yes: sample mean=90
sample variance=25

- P(X|Class=No) = P(Refund=No|Class=No) $\times$ P(Married| Class=No) $\times$ P(Income=120K| Class=No) = 4/7 $\times$ 4/7 $\times$ 0.0072 = 0.0024
- P(X|Class=Yes) = P(Refund=No| Class=Yes) $\times$ P(Married| Class=Yes) $\times$ P(Income=120K| Class=Yes) = 1 $\times$ 0 $\times$ 1.2 $\times$ 10^{-9} = 0

Since P(X|No)P(No) > P(X|Yes)P(Yes)

Therefore P(No|X) > P(Yes|X)
=> Class = No