



Laplacian Correction

Mengatasi Probabilitas dengan nilai 0 untuk
Klasifikasi dengan Naïve Bayes

Margaretha Sulistyoningsih, Ph.D

Data Latih yang sama [1,2]

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no

Kasus dengan Probabilitas 0



- Misalnya data latih adalah data Pembelian Komputer tadi.
- Data Testing adalah Mr.Y:
 - Umur: 31..40.
 - Income: Medium
 - Student: Yes
 - Credit Rating: Excellent
- Tentukan apakah Mr. Y membeli Komputer atau tidak!

Kasus dengan Probabilitas 0



- Karena data latih dengan Umur: 31..40 semua membeli komputer, **maka probabilitas age: 31..40 dengan label buys_computer= No adalah 0 (Probabilitas ==0).**
- Untuk mengatasi hal ini, kita gunakan *Laplacian Correction*.

Laplacian Correction

- Ide: Mengganti nilai 0 dengan nilai kecil.
- Maka: 0 ditambahkan dengan nilai yang kecil, misalnya nilai **1**, begitu juga dengan semua pembilang pada perhitungan probabilitas **pada ciri dan klasifikasi yang bersangkutan**.
 - Kita sebut $\alpha == 1$
- Penyebutnya (penyebut **pada ciri dan klasifikasi yang bersangkutan**) tentu juga harus ditambahkan dengan sebuah nilai yaitu:
 - $k * \alpha$.
- Mengapa dikalikan dengan k? Karena penyebut harus besar, agar nilai kecil. Nilai k adalah jumlah nilai pada ciri (*feature*) yang mengalami zero probability tersebut.

Laplacian Correction [3]

- Mari belajar dengan mengerjakan soal dari data latih tadi:

$$P(\text{age} \leq 30 \mid \text{buys_computer=no}) = 3/5$$

$$P(\text{age} \leq 31..40 \mid \text{buys_computer=no}) = 0/5$$

$$P(\text{age} > 40 \mid \text{buys_computer=no}) = 2/5$$

Maka dengan Laplacian correction:

$$P_{\text{LAP}}(\text{age} \leq 30 \mid \text{buys_computer=no}) = (3+1) / (5+(3*1))$$

$$P_{\text{LAP}}(\text{age} \leq 31..40 \mid \text{buys_computer=no}) = (0+1) / (5+(3*1))$$

$$P_{\text{LAP}}(\text{age} > 40 \mid \text{buys_computer=no}) = (2+1) / (5+(3*1))$$

Laplacian Correction



- Nilai **1** adalah α (alpha) , yaitu nilai kecil yang ditambahkan agar tidak ada zero probability (probabilitas dengan nilai 0).
- Nilai **3** adalah k, yang didapat dari jumlah kategori (kategori nilai-nilai yang ada pada ciri “Age”), yaitu:
 - ≤ 30
 - 31..40
 - > 40
- LANJUTKAN dengan perhitungan Naive Bayes pada PPT Naive Bayes yang lalu.

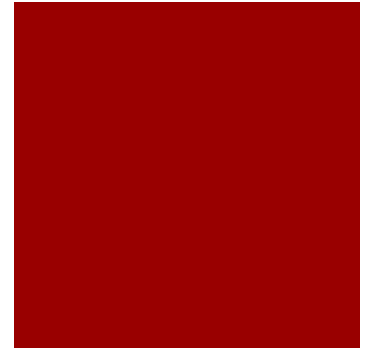
Lanjutkan Perhitungan seperti PPT sebelumnya. Probabilitas dengan nilai 0 diganti dengan hasil Laplacian Correction

$$P(C_i|\mathbf{X}) = P(\mathbf{X}|C_i)P(C_i)$$

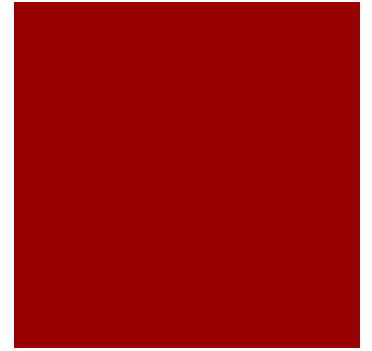
1. $P(C_1) = P(\text{buys_computer} == \text{"yes"}) = \dots$
2. $P(C_2) = P(\text{Buys_computer} == \text{"no"}) = \dots$
3. $P(\text{age} = \text{"31..40"} \mid P(\text{buys_computer} == \text{"yes"}) = \dots$
4. $P(\text{age} = \text{"31..40"} \mid P(\text{buys_computer} == \text{"no"}) = \dots$ (isi dengan hasil Laplacian Correction pada slide 6)
5. $P(\text{income} = \text{"medium"} \mid P(\text{buys_computer} == \text{"yes"}) = \dots$
6. $P(\text{income} = \text{"medium"} \mid P(\text{buys_computer} == \text{"no"}) = \dots$
7. $P(\text{student} = \text{"yes"} \mid P(\text{buys_computer} == \text{"yes"}) = \dots$
8. $P(\text{student} = \text{"yes"} \mid P(\text{buys_computer} == \text{"no"}) = \dots$
9. $P(\text{credit rating} = \text{"excellent"} \mid P(\text{buys_computer} == \text{"yes"}) = \dots$
10. $P(\text{credit rating} = \text{"excellent"} \mid P(\text{buys_computer} == \text{"no"}) = \dots$
11. $P(\mathbf{X} \mid C_1) = \dots$ (semua yang $P(\text{buys_computer} == \text{"yes"})$ dikalikan)
12. $P(\mathbf{X} \mid C_2) = \dots$ (semua yang $P(\text{buys_computer} == \text{"no"})$ dikalikan)
13. $P(\mathbf{X} \mid C_1) * P(C_1) = \dots$ (hasil nomor (11) dikalikan dengan hasil nomor (1))
14. $P(\mathbf{X} \mid C_2) * P(C_2) = \dots$ (hasil nomor (12) dikalikan dengan hasil nomor (2))
15. Kesimpulan: Dengan melihat hasil no (13) dan (14), maka simpulkan apakah Mr. Y akan membeli komputer?

TUGAS LATIHAN MANDIRI

- **Lanjutkan** semua perhitungan dan kumpulkan ke SCE



Referensi



1. Jiawei Han, UIUC CS412, Fall 2017
2. Huan Sun, CSE 5243 Intro to Data Mining, Classification (Basic Concepts & Advanced Methods), Ohio State University, undated.
3. Professor Anita Wasilewska, "Bayesian Classification", Stony Brook University, Undated, Accessed from <https://www3.cs.stonybrook.edu/~cse634/19Bayes2.pdf>, on October 18, 2022.