

NAÏVE BAYES

Ali Akbar Septiandri

Universitas Al-Azhar Indonesia

aliakbars@live.com

March 11, 2020

SELAYANG PANDANG

① ULASAN

② NAÏVE BAYES

Klasifikasi Bayesian

Conditional Independence

Kasus Diskrit

Kasus Kontinu

ULASAN

MINGGU LALU...

- Distribusi uniform
- Distribusi normal/Gaussian
- Maximum Likelihood Estimation
- Distribusi Bernoulli, Binomial, dan Beta

Apa itu central limit theorem?

NAÏVE BAYES

Ingat kembali tentang Bayes' rule!

CONTOH

Seorang dokter tahu bahwa meningitis memiliki probabilitas menyebabkan kekakuan leher sekitar 50%. Kasus meningitis sendiri ditemukan dalam 1 dari 50,000 orang. Di sisi lain, probabilitas ditemukannya kasus kekakuan leher adalah $1/20$.

Pertanyaan: Jika seseorang menderita kekakuan leher, berapa peluangnya orang tersebut terkena meningitis?

CONTOH

DIKETAHUI

$$P(s|m) = 0.5$$

$$P(m) = 1/50,000 = 2 \times 10^{-5}$$

$$P(s) = 1/20 = 0.05$$

SOLUSI

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.5 \times 2 \times 10^{-5}}{0.05} = 0.0002$$

KLASIFIKASI BAYESIAN

- Tujuan: fungsi pembelajaran $f(x) \rightarrow y$
- Klasifikasi **probabilistik**: kelas yang paling mungkin jika diberikan hasil **observasinya**, i.e. $\hat{y} = \underset{y}{\operatorname{argmax}} P(y|x)$
- Probabilitas Bayesian dari sebuah kelas:

$$P(y|x) = \frac{P(x|y)P(y)}{\sum_{y'} P(x|y')P(y')}$$

GLOSARIUM

DATA

Data merupakan **kumpulan objek** (*instances*) yang memiliki **atribut-atribut** tertentu

GLOSARIUM

DATA

Data merupakan **kumpulan objek** (*instances*) yang memiliki **atribut-atribut** tertentu

ATRIBUT

Karakteristik dari suatu objek; **variabel** atau **fitur**

KELAS

Penanda kelompok suatu objek; **label**

GLOSARIUM

DATA

Data merupakan **kumpulan objek** (*instances*) yang memiliki **atribut-atribut** tertentu

ATRIBUT

Karakteristik dari suatu objek; **variabel** atau **fitur**

KELAS

Penanda kelompok suatu objek; **label**

OBJEK

Dikenal juga dengan nama **record**, **poin**, **sampel**, **entitas**, atau **instance**

KLASIFIKASI BAYESIAN: KOMPONEN

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)} = \frac{P(x|y)P(y)}{\sum_{y'} P(x|y')P(y')}$$

- $P(y)$: probabilitas *prior* dari kelas
“mana kelas yang sering muncul, mana yang jarang”
- $P(x|y)$: *class-conditional model* atau *likelihood*
“seberapa sering observasi x dalam kasus y ”
- $P(x)$: faktor normalisasi

ASUMSI INDEPENDENSI

- Apa yang terjadi kalau harus menghitung $P(\mathbf{x}|y)$ sedangkan variabelnya bisa banyak sekali?

ASUMSI INDEPENDENSI

- Apa yang terjadi kalau harus menghitung $P(\mathbf{x}|y)$ sedangkan variabelnya bisa banyak sekali?
- Contoh: MNIST punya 784 variabel, dengan nilai biner saja artinya ada 2^{784} kemungkinan pola!

ASUMSI INDEPENDENSI

- Apa yang terjadi kalau harus menghitung $P(\mathbf{x}|y)$ sedangkan variabelnya bisa banyak sekali?
- Contoh: MNIST punya 784 variabel, dengan nilai biner saja artinya ada 2^{784} kemungkinan pola!
- Namun, kita mengetahui observasi untuk masing-masing nilai x_i untuk setiap kelas

ASUMSI INDEPENDENSI

- Apa yang terjadi kalau harus menghitung $P(\mathbf{x}|y)$ sedangkan variabelnya bisa banyak sekali?
- Contoh: MNIST punya 784 variabel, dengan nilai biner saja artinya ada 2^{784} kemungkinan pola!
- Namun, kita mengetahui observasi untuk **masing-masing nilai** x_i untuk setiap kelas
- Asumsi yang digunakan Naïve Bayes adalah $x_1 \dots x_d$ ***conditionally independent*** jika diberikan y

$$P(\mathbf{x}|y) = \prod_{i=1}^d P(x_i|x_1, \dots, x_{i-1}, y) = \prod_{i=1}^d P(x_i|y)$$

CONDITIONAL INDEPENDENCE

- Probabilitas pergi ke pantai dan *heatstroke* tidak independen

CONDITIONAL INDEPENDENCE

- Probabilitas pergi ke pantai dan *heatstroke* tidak independen
- Bisa jadi independen jika kita tahu cuaca sedang terik

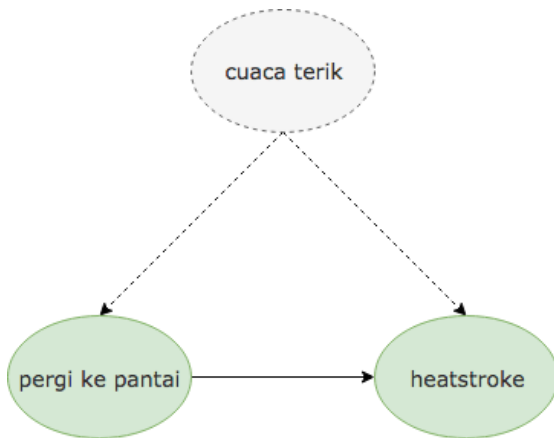
CONDITIONAL INDEPENDENCE

- Probabilitas pergi ke pantai dan *heatstroke* tidak independen
- Bisa jadi independen jika kita tahu cuaca sedang terik
- Cuaca terik “menjelaskan” dependensi antara pergi ke pantai dan *heatstroke*

CONDITIONAL INDEPENDENCE

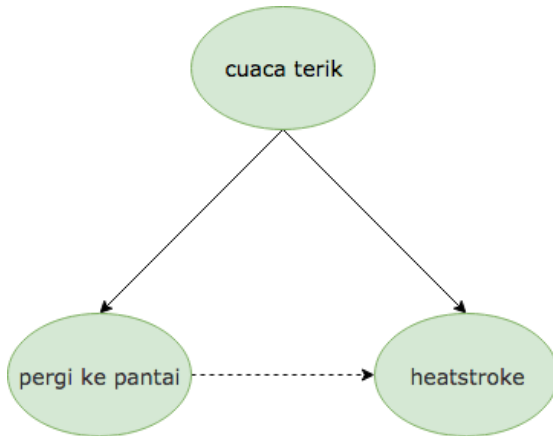
- Probabilitas pergi ke pantai dan *heatstroke* tidak independen
- Bisa jadi independen jika kita tahu cuaca sedang terik
- Cuaca terik “menjelaskan” dependensi antara pergi ke pantai dan *heatstroke*
- Dalam klasifikasi, nilai kelas menjelaskan hubungan antaratribut

ILUSTRASI



GAMBAR: Kita belum mengobservasi cuaca

ILUSTRASI



GAMBAR: Setelah kita tahu kondisi cuaca, *heatstroke* tidak lagi dijelaskan oleh pergi ke pantai

MODEL GENERATIF

- Naïve Bayes menghitung probabilitas untuk masing-masing kelas yang ada

MODEL GENERATIF

- Naïve Bayes menghitung probabilitas untuk masing-masing kelas yang ada
- “Apakah datanya lebih besar probabilitasnya sebagai kelas 1 atau kelas 0?”

MODEL GENERATIF

- Naïve Bayes menghitung probabilitas untuk masing-masing kelas yang ada
- “Apakah datanya lebih besar probabilitasnya sebagai kelas 1 atau kelas 0?”
- Model generatif selalu melakukan klasifikasi probabilistik

MODEL GENERATIF

- Naïve Bayes menghitung probabilitas untuk masing-masing kelas yang ada
- “Apakah datanya lebih besar probabilitasnya sebagai kelas 1 atau kelas 0?”
- Model generatif selalu melakukan klasifikasi probabilistik
- Klasifikasi probabilistik tidak berarti generatif, e.g. *logistic regression*

CONTOH KASUS DISKRIT

Asumsi: Distribusi Bernoulli

Contoh pada kasus identifikasi spam e-mail

D1: “send us your password” (s)

D2: “send us your review” (h)

D3: “review your password” (h)

D4: “review us” (s)

D5: “send your password” (s)

D6: “send us your account” (s)

Dokumen baru: “review us now”

word	spam	ham
password	2/4	1/2
review	1/4	2/2
send	3/4	1/2
us	3/4	1/2
your	3/4	1/2
account	1/4	0/2

$$P(spam) = 4/6, P(ham) = 2/6$$

CONTOH KASUS DISKRIT

$$P(\text{review us}|\text{spam}) = P(0, 1, 0, 1, 0, 0|\text{spam})$$

$$P(\text{review us}|\text{ham}) = P(0, 1, 0, 1, 0, 0|\text{ham})$$

$$P(\text{ham}|\text{review us}) \approx 0.87$$

MASALAH ZERO-FREQUENCY

- Berdasarkan contoh sebelumnya, setiap e-mail dengan kata “account” akan dianggap spam karena
 $P(account|ham) = 0/2$

MASALAH ZERO-FREQUENCY

- Berdasarkan contoh sebelumnya, setiap e-mail dengan kata “account” akan dianggap spam karena $P(account|ham) = 0/2$
- Solusi: Laplace smoothing, i.e. penambahan angka positif kecil ke semua pencacahan

$$P(w|c) = \frac{num(w, c) + \epsilon}{num(c) + 2\epsilon}$$

MASALAH ZERO-FREQUENCY

- Berdasarkan contoh sebelumnya, setiap e-mail dengan kata “account” akan dianggap spam karena
 $P(account|ham) = 0/2$
- Solusi: Laplace smoothing, i.e. penambahan angka positif kecil ke semua pencacahan

$$P(w|c) = \frac{num(w, c) + \epsilon}{num(c) + 2\epsilon}$$

- Nilai ϵ contohnya 1 atau 0.5, tetapi bisa juga dengan $num(w)/num$

MASALAH ZERO-FREQUENCY

- Berdasarkan contoh sebelumnya, setiap e-mail dengan kata “account” akan dianggap spam karena
 $P(account|ham) = 0/2$
- Solusi: Laplace smoothing, i.e. penambahan angka positif kecil ke semua pencacahan

$$P(w|c) = \frac{num(w, c) + \epsilon}{num(c) + 2\epsilon}$$

- Nilai ϵ contohnya 1 atau 0.5, tetapi bisa juga dengan $num(w)/num$
- Kasus ini sering terjadi karena Zipf's law (50% kata hanya muncul sekali)

MISSING VALUES

- Misalkan kita tidak punya nilai untuk atribut X_i , bagaimana kita bisa menghitung $P(X_1 = x_1, \dots, X_i = ?, \dots, X_d = x_d | y)$?

MISSING VALUES

- Misalkan kita tidak punya nilai untuk atribut X_i , bagaimana kita bisa menghitung $P(X_1 = x_1, \dots, X_i = ?, \dots, X_d = x_d | y)$?
- Naïve Bayes dapat mengabaikan atribut tersebut karena *conditional independence*

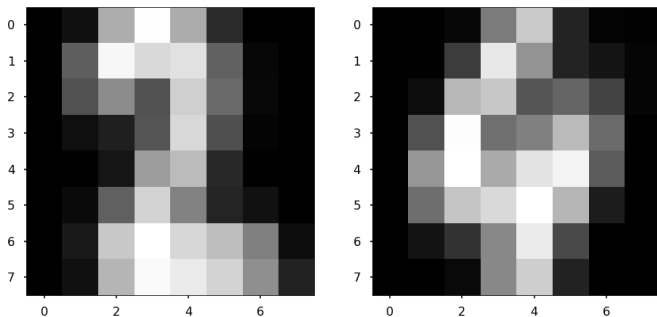
MISSING VALUES

- Misalkan kita tidak punya nilai untuk atribut X_i , bagaimana kita bisa menghitung $P(X_1 = x_1, \dots, X_i = ?, \dots, X_d = x_d | y)$?
- Naïve Bayes dapat mengabaikan atribut tersebut karena *conditional independence*
- Hitung saja berdasarkan atribut yang bernilai!

MISSING VALUES

- Misalkan kita tidak punya nilai untuk atribut X_i , bagaimana kita bisa menghitung $P(X_1 = x_1, \dots, X_i = ?, \dots, X_d = x_d | y)$?
- Naïve Bayes dapat mengabaikan atribut tersebut karena *conditional independence*
- Hitung saja berdasarkan atribut yang bernilai!
- Nilai yang hilang tersebut tidak perlu diganti

CONTOH KASUS KONTINU



GAMBAR: Gambar rata-rata pixel angka 2 vs 4. Apakah kita dapat menggunakan pixel (0,2) dan (4,6) saja untuk mengklasifikasikan gambar?

CONTOH KASUS KONTINU

- Identifikasi angka 2 atau 4 berdasarkan pixel $(0,2)$ dan $(4,6)$: $y = \{2, 4\}$, atribut: $\{tl, mr\}$

CONTOH KASUS KONTINU

- Identifikasi **angka 2 atau 4** berdasarkan pixel (0,2) dan (4,6): $y = \{2, 4\}$, atribut: $\{tl, mr\}$
- Probabilitas kelas:
$$P(y = 2) = \frac{177}{177+181} \approx 0.49,$$
$$P(y = 4) = 1 - P(y = 2) \approx 0.51$$

CONTOH KASUS KONTINU

- Identifikasi **angka 2 atau 4** berdasarkan pixel (0,2) dan (4,6): $y = \{2, 4\}$, atribut: $\{tl, mr\}$
- Probabilitas kelas:
$$P(y = 2) = \frac{177}{177+181} \approx 0.49,$$
$$P(y = 4) = 1 - P(y = 2) \approx 0.51$$
- Asumsi: atribut terdistribusi Gaussian dan independen jika diketahui kelasnya

DISTRIBUSI GAUSSIAN

PDF

$$P(x|\mu, \sigma^2) = \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}$$

MAXIMUM LIKELIHOOD ESTIMATION (MLE)

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu})^2$$

PARAMETER DISTRIBUSI GAUSSIAN UNTUK NAÏVE BAYES

Dicocokkan dengan *maximum likelihood estimation* (MLE)
untuk Gaussian, e.g.

$$\hat{\mu}_{tl,2} = \frac{1}{177} \sum_{i=1}^{177} tl_i$$

$$\hat{\sigma}_{tl,2}^2 = \frac{1}{177} \sum_{i=1}^{177} (tl_i - \hat{\mu}_{tl,2})^2$$

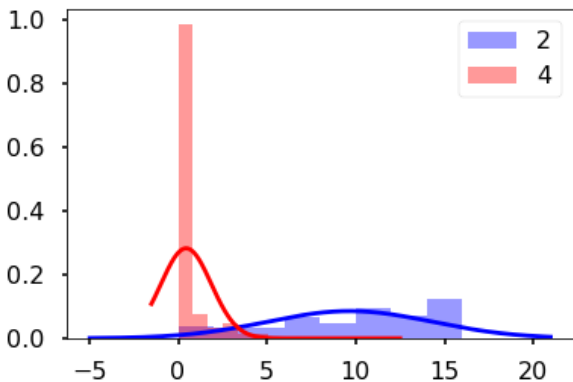
NAÏVE BAYES

$$\begin{aligned} P(y|\mathbf{x}) &= \frac{P(\mathbf{x}|y)P(y)}{P(\mathbf{x})} \\ &= \frac{P(x_{tl}|y)P(x_{mr}|y)P(y)}{\sum_y P(x_{tl}|y)P(x_{mr}|y)P(y)} \end{aligned}$$

BATAS KEPUTUSAN

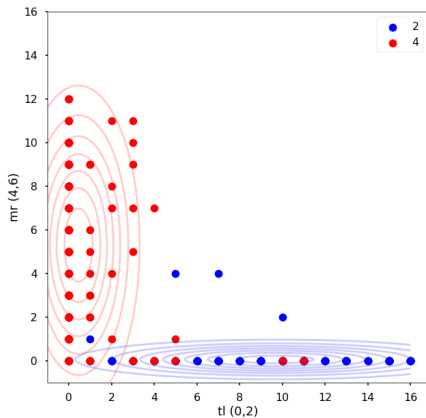
- Beda rataan, variansi sama: garis lurus atau bidang lurus
- Rataan sama, beda variansi: lingkaran atau elips
- Kasus umum: kurva parabola

BATAS KEPUTUSAN



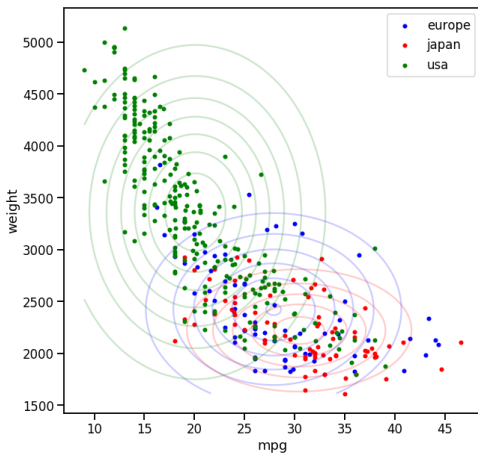
GAMBAR: Dalam satu dimensi, perbandingan pixel (0,2) dari dua kelas

BATAS KEPUTUSAN



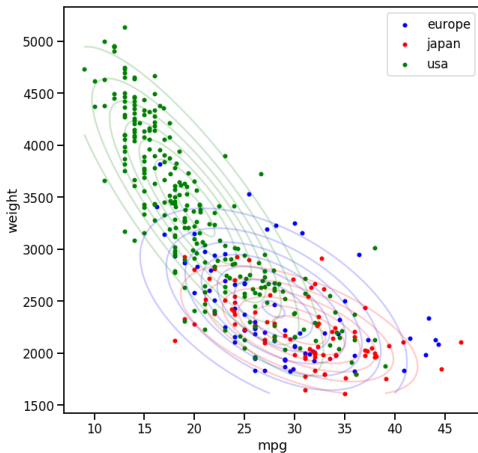
GAMBAR: Dalam dua dimensi, perbandingan pixel $(0,2)$ dan $(4,6)$ dari dua kelas

NAÏVE BAYES PADA DUA DIMENSI



GAMBAR: Asumsi naif \rightarrow *diagonal covariance matrix*

MELUNAKKAN ASUMSI



GAMBAR: Menggunakan *full covariance matrix* untuk melunakkan asumsi

IKHTISAR

- Naïve Bayes
- Conditional independence
- Penggunaan distribusi Gaussian dan Bernoulli untuk NB
- Diagonal dan full covariance matrix saat klasifikasi

PERTEMUAN BERIKUTNYA

- Dimensionality Reduction
- Eigenvector & Eigenvalue
- Principal Component Analysis

Salindia ini dipersiapkan dengan sangat dipengaruhi oleh:
Victor Lavrenko (2014)

Terima kasih