

# Ulasan Probabilitas

---

Ali Akbar Septiandri

November 11, 2017

untuk Astra Graphia IT

1. Pendahuluan
2. Bayes' Rule
3. Ekspektasi, Variansi, dan Kovariansi
4. Distribusi Gaussian
5. Distribusi Bernoulli

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. (Part I.3 Probability and Information Theory)
2. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann. (Chapter 4. Algorithms: the basic method)
3. Murphy, K. P. (2012). Machine learning: a probabilistic perspective. MIT Press. (Chapter 3. Generative models for discrete data)

# Pendahuluan

---

# Mengapa probabilitas?

Ada tiga sumber ketidakpastian:

1. Sistem yang stokastik secara inheren, e.g. mekanika kuantum
2. Tidak semua dapat diobservasi
3. Pemodelan yang tidak lengkap, i.e. keterbatasan algoritma

# Peubah Acak

- Dalam pendekatan probabilistik, data dapat dilihat sebagai observasi yang muncul dari model probabilitas untuk sebuah **peubah acak** (*random variable* atau r.v.)

# Peubah Acak

- Dalam pendekatan probabilistik, data dapat dilihat sebagai observasi yang muncul dari model probabilitas untuk sebuah **peubah acak** (*random variable* atau r.v.)
- Jika diberikan peubah acak diskrit  $A$ , maka  $P(A)$  adalah fungsi yang memetakan kemungkinan munculnya kelas, kategori, atau kondisi dari  $A$ , atau dikenal sebagai **probability mass function** (PMF)

# Peubah Acak

- Dalam pendekatan probabilistik, data dapat dilihat sebagai observasi yang muncul dari model probabilitas untuk sebuah **peubah acak** (*random variable* atau r.v.)
- Jika diberikan peubah acak diskrit  $A$ , maka  $P(A)$  adalah fungsi yang memetakan kemungkinan munculnya kelas, kategori, atau kondisi dari  $A$ , atau dikenal sebagai **probability mass function** (PMF)
- Jika diberikan peubah acak kontinu  $x$ , maka  $P(x)$  adalah fungsi yang memetakan probabilitas munculnya suatu nilai berdasarkan semua nilai yang ada, atau dikenal sebagai **probability density function** (PDF)



- Terkadang,  $P(A = a)$  dengan  $a$  adalah salah satu kondisi dari r.v.  $A$  disingkat sebagai  $P(a)$  saja
- Demikian halnya dengan  $x_1$  untuk menggambarkan nilai r.v.  $x$  sehingga  $P(x = x_1)$  dapat dituliskan sebagai  $P(x_1)$  saja

# Product Rule & Sum Rule

## Product rule

$$P(A, B) = P(A|B)P(B)$$

# Product Rule & Sum Rule

## Product rule

$$P(A, B) = P(A|B)P(B)$$

## Sum Rule

Jika diberikan *joint probability* variabel  $X_1, X_2, \dots, X_N$ , *marginal probability* dari sebuah variabel bisa didapatkan dengan penjumlahan dari semua variabel yang lainnya.

# Product Rule & Sum Rule

## Product rule

$$P(A, B) = P(A|B)P(B)$$

## Sum Rule

Jika diberikan *joint probability* variabel  $X_1, X_2, \dots, X_N$ , *marginal probability* dari sebuah variabel bisa didapatkan dengan penjumlahan dari semua variabel yang lainnya.

## Sum Rule

$$P(X_1) = \sum_{x_2} \dots \sum_{x_N} P(X_1, X_2 = x_2, \dots, X_N = x_N)$$

- Notasi pada *sum rule* dapat disederhanakan menjadi

$$P(x_1) = \sum_{x_2} \dots \sum_{x_N} P(x_1, x_2, \dots, x_N)$$

- Untuk r.v. kontinu, penjumlahannya diganti dengan integral

$$P(x_1) = \int_{x_2} \dots \int_{x_N} P(x_1, x_2, \dots, x_N) dx_2 \dots dx_N$$

- Prosedur ini dikenal dengan nama **marginalisasi**

# Bayes' Rule

---

# Bayes' Rule

Berdasarkan *product rule*, kita tahu bahwa

$$P(A, C) = P(A|C)P(C)$$

Namun, kita juga bisa melihat bahwa

$$P(A, C) = P(C|A)P(A)$$

sehingga dapat dirumuskan dengan Bayes' rule

$$P(C|A) = \frac{P(A|C)P(C)}{P(A)}$$

## Contoh 1

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 1

### Diketahui

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

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

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

### Solusi

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

## Contoh 2

Terdapat dua tim sepakbola, tim 0 dan tim 1. Tim 0 memenangkan 65% pertandingan dalam pertemuan kedua tim tersebut, sedangkan tim 1 memenangkan sisanya. Dari semua kemenangan tim 0, hanya 30% terjadi saat keduanya bertanding di kandang tim 1. Di sisi lain, 75% kemenangan tim 1 terjadi saat mereka bermain di kandang.

**Pertanyaan:** Berapa peluang tim 1 akan menang jika di pertandingan berikutnya mereka akan bermain di kandang?

# Ekspektasi, Variansi, dan Kovariansi

---

**Ekspektasi** atau **nilai harapan** dari sebuah fungsi  $f(x)$  dengan probabilitas  $P(x)$  adalah rata-rata dari  $f$  saat  $x$  diambil dari  $P$ .

**Peubah acak diskrit**

$$\mathbb{E}[f(x)] = \sum_x P(x)f(x)$$

**Peubah acak kontinu**

$$\mathbb{E}[f(x)] = \int p(x)f(x)dx$$

Nilai harapan, ekspektasi, atau *expected value* bersifat linear, contohnya

$$\mathbb{E}_x[\alpha f(x) + \beta g(x)] = \alpha \mathbb{E}_x[f(x)] + \beta \mathbb{E}_x[g(x)]$$

saat  $\alpha$  dan  $\beta$  independen terhadap  $x$ .

**Variansi** menggambarkan seberapa besar perubahan dari fungsi untuk peubah acak  $x$  saat kita mengambil sampel dari distribusinya.

$$\begin{aligned} \text{Var}(f(x)) &= \mathbb{E} [(f(x) - \mathbb{E}[f(x)])^2] \\ &= \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2 \end{aligned}$$

Akar dari variansi dikenal sebagai **simpangan baku**.



**Kovariansi** menggambarkan hubungan linear antara dua variabel, dan skala dari masing-masing variabel tersebut.

$$\text{Cov}(f(x), g(y)) = \mathbb{E}[(f(x) - \mathbb{E}[f(x)])(g(y) - \mathbb{E}[g(y)])]$$

# Distribusi Gaussian

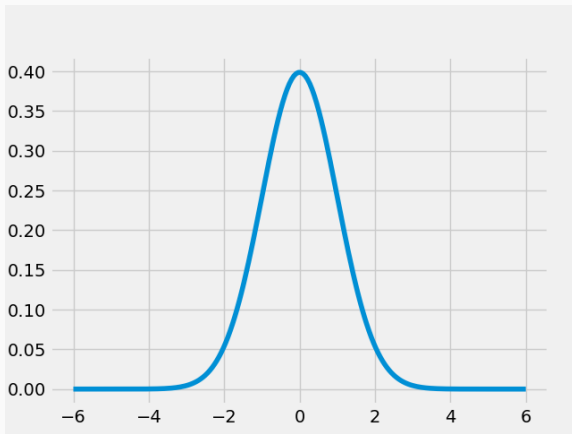
---

# Distribusi Gaussian/Normal

- Salah satu distribusi yang paling umum dalam kasus peubah acak kontinu
- Ada hubungannya dengan *central limit theorem*
- Distribusi Gaussian satu dimensi didefinisikan sebagai

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

- $x \sim \mathcal{N}(\mu, \sigma^2)$  ( $x$  terdistribusi ...)
- $\mu$  adalah rata-rata dari Gaussian dan  $\sigma^2$  adalah variansinya



**Gambar 1:** Distribusi Gaussian untuk satu variabel

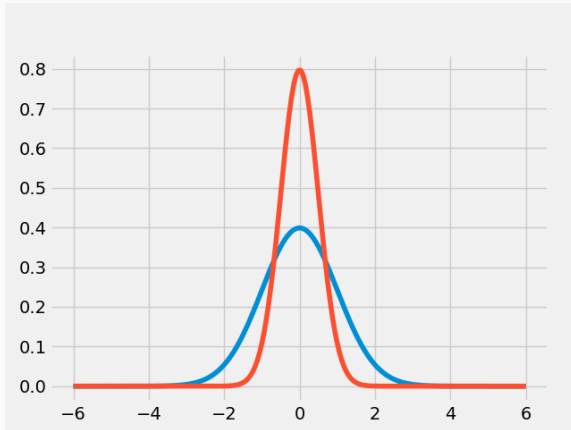
- Ingat bahwa semua distribusi integrasinya harus bernilai satu

- Ingat bahwa semua distribusi integrasinya harus bernilai satu
- $\sqrt{2\pi\sigma^2}$  disebut sebagai **konstanta normalisasi** untuk memastikan hal tersebut

- Ingat bahwa semua distribusi integrasinya harus bernilai satu
- $\sqrt{2\pi\sigma^2}$  disebut sebagai **konstanta normalisasi** untuk memastikan hal tersebut
- Jadi, distribusi Gaussian dengan **variansi yang lebih rendah** akan memiliki puncak yang lebih tinggi



# Normalisasi



**Gambar 2:** Dua distribusi Gaussian dengan variansi yang berbeda

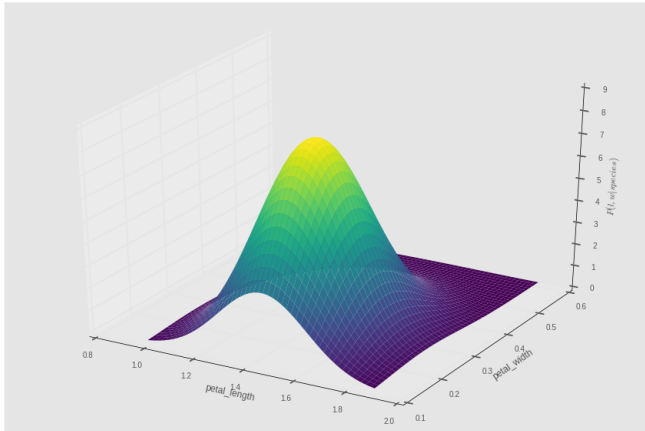
- Vektor  $\mathbf{x}$  adalah Gaussian multivariat jika untuk *mean*  $\mu$  dan matriks kovariansi  $\Sigma$  distribusinya

$$p(\mathbf{x}|\mu, \Sigma) = \frac{1}{|(2\pi)\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right)$$

- $\Sigma$  disebut sebagai matriks kovariansi, i.e. setiap elemennya adalah  $\sigma_{ij} = \text{Cov}(X_i, X_j)$ , dengan

$$\text{Cov}(X_i, X_j) = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)]$$

- $\Sigma$  harus simetris



**Gambar 3:** Distribusi Gaussian untuk dua variabel (*bivariate Gaussian*)

- Penjumlahan dari Gaussian RVs bersifat Gaussian
- Model linear dari Gaussian juga bersifat Gaussian untuk distribusi gabungannya
- Jika  $p(\mathbf{x}, \mathbf{y})$  adalah Gaussian multivariat, maka  $p(\mathbf{x})$ ,  $p(\mathbf{y})$ , dan  $p(\mathbf{x}|\mathbf{y})$ ,  $p(\mathbf{y}|\mathbf{x})$  juga bersifat Gaussian

# Distribusi Bernoulli

---

# Distribusi Bernoulli

- Distribusi Bernoulli adalah distribusi untuk peubah acak yang bernilai biner
- Hanya punya satu parameter  $\theta$
- Didefinisikan sebagai

$$p(X|\theta) = \begin{cases} \theta & , X = 1 \\ 1 - \theta & , X = 0 \end{cases}$$

atau dapat ditulis sebagai

$$p(x) = \theta^x(1 - \theta)^{1-x}$$

- Properti

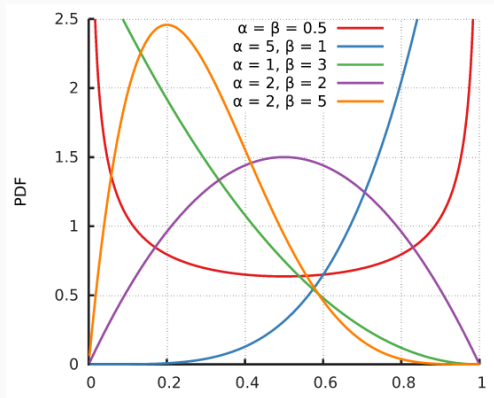
$$\mathbb{E}[x] = \theta$$

$$\text{Var}(x) = \theta(1 - \theta)$$

Lihat juga tentang:

- distribusi Multinoulli
- distribusi Binomial
- distribusi Beta

# Distribusi Beta



**Gambar 4:** Distribusi Beta yang diatur oleh dua parameter,  $\alpha$  dan  $\beta$



Terima kasih