

Pengenalan Pola

Pendahuluan

Ali Akbar Septiandri

Universitas Al-Azhar Indonesia

aliakbars@live.com

March 5, 2018

Selayang Pandang

① Administrasi

Tentang Perkuliahan
Referensi

② Pembelajaran Mesin & Pengenalan Pola Definisi

③ Tugas dalam Machine Learning

Klasifikasi

Regresi

Supervised Learning

Clustering

Representasi Data

Administrasi

Mata Kuliah Terkait

Prerequisites

- Statistika & Probabilitas, e.g. distribusi normal, ekspektasi
- Aljabar Linear, e.g. vektor dan nilai eigen, perkalian matriks
- Kalkulus, e.g. diferensial, integral
- Kecerdasan Buatan

Paralel/Saran/Lanjutan

- Pemrograman Python
- Data/Text Mining

Aturan Perkuliahan

- Dosen pengampu: Ali Akbar S. dan Firmansyah Asnawi
- Materi bisa dilihat di <http://uai.aliakbars.com/mlpr/>
- Kuliah setiap hari Selasa, pukul 07.00-09.30
- Office hours: mulai minggu depan, hari Kamis, pukul 08.00-09.00
- Bahasa/teknologi pengantar: Python, PyLab, Jupyter Notebook

Aturan Perkuliahan

- Terdapat 4 tugas
- Kuis yang tidak masuk komponen penilaian
- Ujian Tengah Semester dan makalah pengganti UAS (tidak ada ujian perbaikan)
- Komponen nilai: 40% tugas, 30% UTS, 30% makalah dan poster

Aturan dalam tugas

- Secara *default*, setiap tugas bersifat **individual**
- Silakan berdiskusi, tapi **jangan menyalin kode atau tulisan teman**
- Toleransi **total 3 hari** keterlambatan tugas, bukan pada makalah dan poster
- **Keterlambatan pengumpulan** akan berakibat pada pengurangan nilai
- Pengumpulan tugas dilakukan melalui situs **e-learning**

Percayalah, tugas itu berpengaruh besar ke nilai akhir.

Referensi

Buku dan materi daring yang bisa dijadikan referensi:

- 1 VanderPlas, J. (2016). *Python Data Science Handbook*. O'Reilly Media. (tersedia online)
- 2 Bishop, C.M. (2007). *Pattern Recognition and Machine Learning*. Springer.
- 3 Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78-87. (tersedia online)

Referensi

Beberapa kuliah terkait:

- 1 University of Edinburgh: Introductory Applied Machine Learning
- 2 University of Edinburgh: Machine Learning and Pattern Recognition (post-grad level)
- 3 Stanford CS229: Machine Learning
- 4 University of Oxford: Machine Learning
- 5 Caltech: Machine Learning

Materi

Sebelum UTS

- 1 Intro to MLPR
- 2 Probability
- 3 Some useful distributions
- 4 Naïve Bayes & LDA
- 5 PCA
- 6 Linear & logistic regression
- 7 Model evaluation

Setelah UTS

- 1 Neural Networks (Part 1)
- 2 Neural Networks (Part 2)
- 3 k-Nearest Neighbours
- 4 k-Means clustering
- 5 Gaussian Mixture Models
- 6 Bimbingan makalah
- 7 Kuliah tamu

Pembelajaran Mesin & Pengenalan Pola

Apa itu Machine Learning?

- Menemukan pola dalam data dan menggunakannya untuk melakukan prediksi
- Bagaimana cara menyelesaikan dengan menggunakan komputer?
 - Masalahnya, kita tidak tahu cara menulis programnya
 - ...tapi kita punya contoh data

Mengapa Machine Learning?

- Data ada di mana-mana
- ML menggabungkan teori dan praktik
- Sudah berhasil menyelesaikan banyak kasus AI

Apa hubungannya dengan pengenalan pola?

Pengenalan Pola

- Pengenalan pola, pembelajaran mesin, *data mining*, dan *knowledge discovery in databases* (KDD) terkadang sulit dibedakan
- Pengenalan pola awalnya dikenal dalam tugas-tugas *computer vision*
- *Pattern recognition* \neq *pattern matching*

Tugas dalam Machine Learning

Klasifikasi

- 1 Memprediksi nilai yang sudah pasti

Klasifikasi

- 1 Memprediksi nilai yang sudah pasti
- 2 *Biasanya* direpresentasikan sebagai kelas biner $\{0, 1\}$ atau $\{-1, 1\}$

Klasifikasi

- 1 Memprediksi nilai yang sudah pasti
- 2 *Biasanya* direpresentasikan sebagai kelas biner $\{0, 1\}$ atau $\{-1, 1\}$
- 3 Membutuhkan label

Klasifikasi

- 1 Memprediksi nilai yang sudah pasti
- 2 *Biasanya* direpresentasikan sebagai kelas biner $\{0, 1\}$ atau $\{-1, 1\}$
- 3 Membutuhkan label
- 4 Mempunyai *evaluation metrics* yang jelas, e.g. akurasi

Klasifikasi

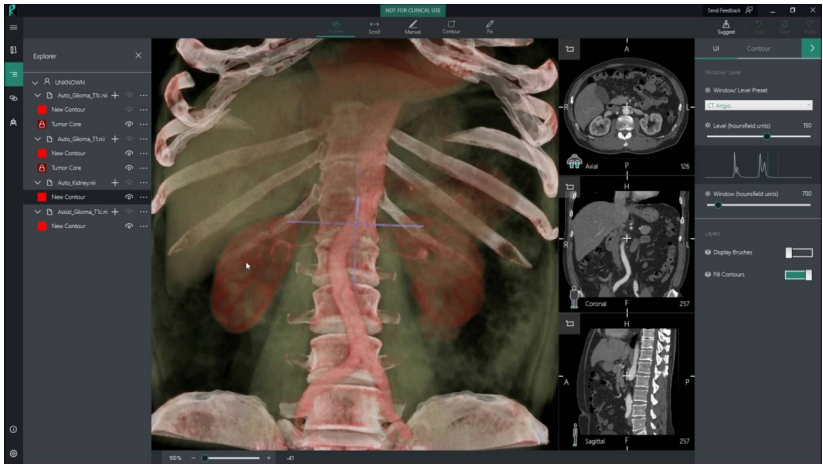
- 1 Memprediksi nilai yang sudah pasti
- 2 *Biasanya* direpresentasikan sebagai kelas biner $\{0, 1\}$ atau $\{-1, 1\}$
- 3 Membutuhkan label
- 4 Mempunyai *evaluation metrics* yang jelas, e.g. akurasi
- 5 Contoh: identifikasi spam, MNIST digit recognition

Quick, Draw!



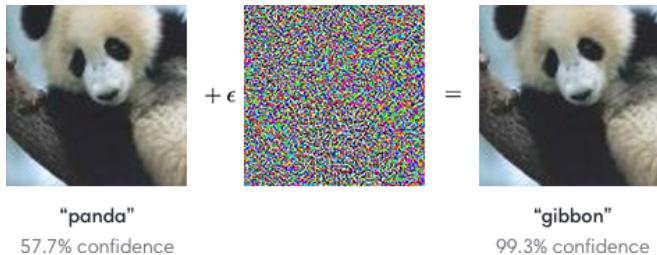
Gambar: Quick, Draw! The Data dari Google

InnerEye Research Project



Gambar: Project InnerEye untuk membantu dokter mendeteksi penyakit pada organ dalam dari Microsoft

Misklasifikasi dalam Pengenalan Objek



Gambar: Kesalahan model *machine learning* akibat *adversarial examples*

Regresi

- 1 Membutuhkan label

Regresi

- 1 Membutuhkan label
- 2 Memprediksi nilai kontinu

Regresi

- 1 Membutuhkan label
- 2 Memprediksi nilai kontinu
- 3 *Evaluation metrics* berupa *error*, e.g. Mean Squared Error (MSE), Mean Absolute Error (MAE)

Regresi

- 1 Membutuhkan label
- 2 Memprediksi nilai kontinu
- 3 *Evaluation metrics* berupa *error*, e.g. Mean Squared Error (MSE), Mean Absolute Error (MAE)
- 4 Contoh: prediksi nilai saham, jumlah RT dari suatu *tweet*

Inpainting



(a) Input context



(b) Human artist



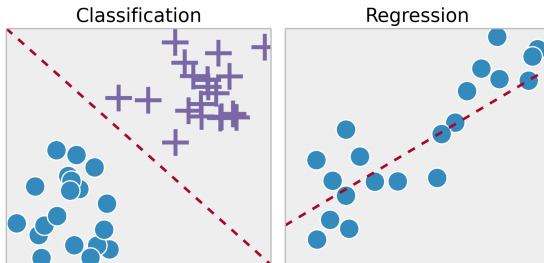
(c) Context Encoder
(L_2 loss)



(d) Context Encoder
(L_2 + Adversarial loss)

Gambar: Mengisi potongan gambar yang hilang [Pathak, 2016]

Klasifikasi vs Regresi



Gambar: Perbedaan klasifikasi dan regresi [Rossant, 2014]

Klasifikasi dan Regresi

Fungsi

Kedua tugas ini dapat dilihat sebagai fungsi f yang memetakan atribut x ke label y .

Probabilitas dalam Machine Learning

- Asumsikan Anda diberi kasus berupa klasifikasi artikel berita. Anggap x adalah dokumen, dan y adalah label.
 $y \in \{ \text{"Olahraga"}, \text{"Politik"} \}$
- Anda diminta membuat fungsi f dalam Java yang menerima masukan berupa x dan mengeluarkan y
- Jika saya akan membayar Anda Rp 1000 setiap artikel politik yang benar, dan Rp 1 juta untuk setiap artikel olahraga yang Anda dapat temukan dengan benar, bagaimana Anda akan membuat f ?

Probabilitas dalam Machine Learning

- Agar lebih sulit, bagaimana kalau Anda akan didenda Rp 10.000 setiap ada dokumen yang salah diklasifikasi? Apa yang Anda akan lakukan?

Probabilitas dalam Machine Learning

- Agar lebih sulit, bagaimana kalau Anda akan didenda Rp 10.000 setiap ada dokumen yang salah diklasifikasi? Apa yang Anda akan lakukan?
- Jawabannya: **Jangan buat fungsi**. Definisikan probabilitas $p(y|\mathbf{x})$. Lalu, ambil keputusan yang dapat memaksimalkan keuntungan Anda.

“No free lunch”

Clustering

- 1 Mencoba memberikan deskripsi terhadap data

Clustering

- ① Mencoba memberikan deskripsi terhadap data
- ② Tidak berhubungan dengan label

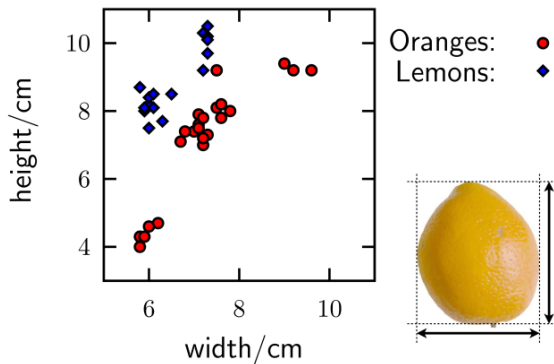
Clustering

- ① Mencoba memberikan deskripsi terhadap data
- ② Tidak berhubungan dengan label
- ③ Menemukan pola yang “menarik” dalam data

Clustering

- ① Mencoba memberikan deskripsi terhadap data
- ② Tidak berhubungan dengan label
- ③ Menemukan pola yang “menarik” dalam data
- ④ Tidak mempunyai *evaluation metrics* yang pasti

Contoh Clustering



Gambar: *Clustering* buah lemon dan jeruk [Murray, 2011]

Bagaimana merepresentasikan datanya?

Representasi Data

- dokumen?

Representasi Data

- dokumen?
- gambar?

Representasi Data

- dokumen?
- gambar?
- video?

Representasi Data

- dokumen?
- gambar?
- video?
- suara?

Representasi Data

- dokumen? *bag of words, word embedding*
- gambar? *pixels, regions*
- video? *pixels in frames*
- suara? *MFCC*

Ikhtisar

- ML & PR ada di mana-mana dan berguna
- Butuh pemahaman teoretis untuk dapat memahami algoritma dan membuat yang baru
- “No free lunch”
- Buat model, bukan algoritma
- Mungkin butuh pandangan probabilistik

Pertemuan Berikutnya

- Probabilitas
- Peubah acak
- Ekspektasi
- Peluang bersyarat
- Bayes' rule
- Beberapa distribusi

Referensi



Ian J. Goodfellow et al. (2015)

Explaining and Harnessing Adversarial Examples
ICLR 2015



Cyrille Rossant (2014)

Introduction to Machine Learning in Python with scikit-learn
<http://ipython-books.github.io/featured-04/>



Deepak Pathak et al. (2016)

Context Encoders: Feature Learning by Inpainting
CVPR 2016

Referensi



Iain Murray (2011)

Oranges, Lemons and Apples dataset

http://homepages.inf.ed.ac.uk/imurray2/teaching/oranges_and_lemons/

Terima kasih