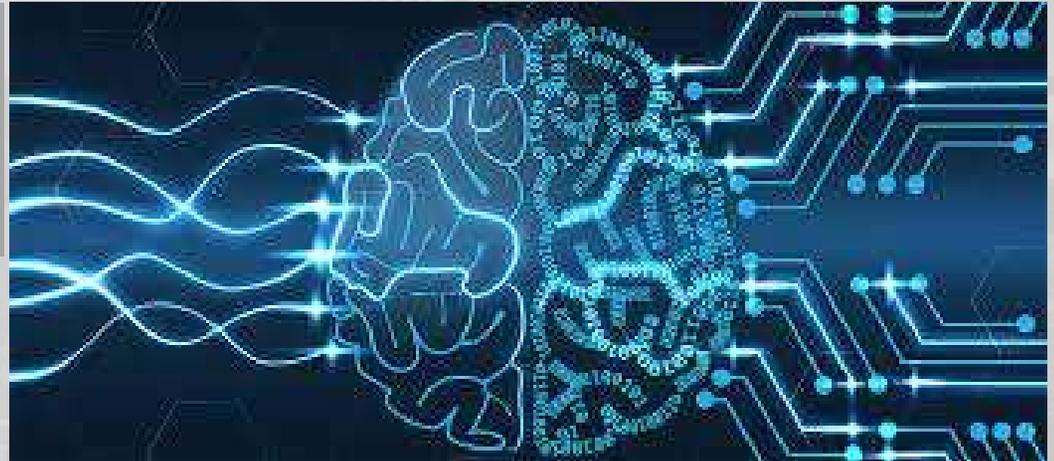


14620323
DEEP LEARNING



Machine Learning Basics



Universitas 17 Agustus 1945 Surabaya

Teknik Informatika

PENGAMPU



Dr. Fajar Astuti Hermawati, S.Kom., M.Kom.



Elsen Ronando, S.Si., M.Si



Bagus Hardiansyah, S.Kom., M.Si



Andrey Kartika Widhy H., S.Kom., M.Kom.



Capaian Pembelajaran

- Mampu mengidentifikasi konsep matematika dan mesin pembelajaran dasar untuk algoritma deep learning. [C2, A3]



Universitas 17 Agustus 1945 Surabaya



Teknik Informatika

Bahan Kajian

- Linear Algebra
- Probability and Information Theory
- Numerical Computation
- **Machine Learning Basics**



Universitas 17 Agustus 1945 Surabaya



Teknik Informatika

Machine Learning Basics



Universitas 17 Agustus 1945 Surabaya

Teknik Informatika

Why?

- Pembelajaran mendalam adalah jenis pembelajaran mesin tertentu.
- Untuk memahami deep learning dengan baik, seseorang harus memiliki pemahaman yang kuat tentang prinsip-prinsip dasar machine learning.



Learning Algorithms

- Algoritma pembelajaran mesin adalah algoritma yang mampu belajar dari data.
- Tapi apa yang kita maksud dengan **belajar**? Mitchell (1997) memberikan definisi ringkas:
 - “Sebuah program komputer dikatakan belajar dari pengalaman E sehubungan dengan beberapa kelas tugas T dan ukuran kinerja P , jika kinerjanya pada tugas di T , yang diukur dengan P , meningkat dengan pengalaman E .”



Task, T

- Dalam definisi kata “**tugas/task**” yang relatif formal ini, **proses pembelajaran itu sendiri bukanlah tugas.**
- Belajar adalah cara kita mencapai kemampuan untuk melakukan tugas.
 - Misalnya, jika kita ingin robot bisa berjalan, maka tugasnya adalah berjalan.
 - Kita dapat memprogram robot untuk belajar berjalan, atau kita dapat mencoba langsung menulis program yang menentukan cara berjalan secara manual.



Task, T

- **Tugas pembelajaran mesin** biasanya dijelaskan dalam hal bagaimana **sistem pembelajaran mesin harus memproses sebuah contoh (example)**.
- **Contoh (example)** adalah **kumpulan fitur** yang telah diukur secara kuantitatif dari beberapa objek atau peristiwa yang ingin diproses oleh sistem pembelajaran mesin.
- Kita biasanya mewakili **contoh** sebagai vektor $\mathbf{x} \in \mathbb{R}^n$ di mana setiap entri x_i vektor adalah fitur lain.
- Sebagai contoh, fitur dari sebuah gambar biasanya merupakan nilai dari piksel pada gambar tersebut.



Task, T

- Banyak jenis tugas dapat diselesaikan dengan pembelajaran mesin. Beberapa tugas pembelajaran mesin yang paling umum meliputi yang berikut:
- **Klasifikasi:**
 - Dalam jenis tugas ini, program komputer diminta untuk menentukan yang mana dari k kategori beberapa masukan.
 - Untuk mengatasi tugas ini, algoritma pembelajaran biasanya diminta untuk menghasilkan fungsi $f: \mathbb{R}^n \rightarrow \{1, \dots, k\}$.
 - Ketika $y = f(x)$, model menetapkan input yang dijelaskan oleh vektor x ke kategori yang diidentifikasi dengan kode numerik y .



Task, T

- **Klasifikasi dengan input yang hilang:**

- Ketika beberapa input mungkin hilang, algoritma pembelajaran harus mempelajari sekumpulan fungsi.
- Setiap fungsi berhubungan dengan mengklasifikasikan x dengan subset berbeda dari inputnya yang hilang.
- Situasi seperti ini sering muncul dalam diagnosis medis karena banyak jenis tes medis yang mahal atau invasif.
- Salah satu cara untuk secara efisien mendefinisikan sekumpulan besar fungsi adalah dengan mempelajari **distribusi probabilitas pada semua variabel yang relevan**, kemudian menyelesaikan tugas klasifikasi dengan meminggirkan variabel yang hilang.



Task, T

- Regresi:
 - Dalam jenis tugas ini, program komputer diminta **untuk memprediksi nilai numerik** dengan memberikan beberapa masukan.
 - Untuk menyelesaikan tugas ini, algoritma pembelajaran diminta untuk mengeluarkan fungsi $f: \mathbb{R}^n \rightarrow \mathbb{R}$.
 - Jenis tugas ini mirip dengan klasifikasi, hanya saja format keluarannya berbeda.
 - Contoh tugas regresi adalah prediksi jumlah klaim yang diharapkan yang akan dibuat oleh orang yang diasuransikan (digunakan untuk menetapkan premi asuransi), atau prediksi harga sekuritas di masa mendatang.



Task, T

- **Transkripsi:**

- Dalam jenis tugas ini, sistem pembelajaran mesin diminta untuk mengamati representasi yang relatif tidak terstruktur dari beberapa jenis data dan mentranskripsikan informasi tersebut ke dalam bentuk tekstual diskrit.
- Misalnya, dalam pengenalan karakter optik, program komputer diperlihatkan foto yang berisi gambar teks dan diminta untuk mengembalikan teks ini dalam bentuk urutan karakter (misalnya, dalam format ASCII atau Unicode).
- Google StreetView menggunakan pembelajaran mendalam untuk memproses nomor alamat dengan cara ini (Goodfellow et al., 2014d).



Task, T

- **Terjemahan mesin:**

- Dalam tugas terjemahan mesin, input sudah terdiri dari urutan simbol dalam beberapa bahasa, dan program komputer harus mengubahnya menjadi urutan simbol dalam bahasa lain.
- Ini umumnya diterapkan pada bahasa alami, seperti menerjemahkan dari bahasa Inggris ke bahasa Prancis.



Task, T

- Output terstruktur:
 - Tugas output terstruktur melibatkan tugas apa pun di mana outputnya adalah vektor (atau struktur data lain yang berisi banyak nilai) dengan hubungan penting antara elemen yang berbeda.
 - Ini adalah kategori yang luas dan mencakup tugas transkripsi dan terjemahan yang dijelaskan di atas, serta banyak tugas lainnya.
 - Salah satu contohnya adalah *parsing* — memetakan kalimat bahasa alami ke dalam pohon yang menggambarkan struktur tata bahasanya dengan menandai simpul pohon sebagai kata kerja, kata benda, kata keterangan, dan sebagainya.



Task, T

- Deteksi anomali:
 - Dalam jenis tugas ini, program komputer menyaring serangkaian peristiwa atau objek dan menandai beberapa di antaranya sebagai tidak biasa atau tidak biasa.
 - Contoh tugas deteksi anomali adalah deteksi penipuan kartu kredit. Dengan memodelkan kebiasaan pembelian, perusahaan kartu kredit dapat mendeteksi penyalahgunaan kartu. Jika seorang pencuri mencuri informasi kartu kredit atau kartu kredit, pembelian pencuri sering kali berasal dari distribusi probabilitas yang berbeda dari tipe pembelian kita.



Task, T

- Sintesis dan pengambilan sampel:
 - Dalam jenis tugas ini, algoritma pembelajaran mesin diminta untuk menghasilkan contoh baru yang serupa dengan yang ada di data pelatihan.
 - Sintesis dan pengambilan sampel melalui pembelajaran mesin dapat berguna untuk aplikasi media ketika membuat konten dalam jumlah besar secara manual akan mahal, membosankan, atau membutuhkan terlalu banyak waktu.
 - Misalnya, video game dapat secara otomatis menghasilkan tekstur untuk objek atau lanskap besar, daripada meminta artis untuk memberi label setiap pi secara manual.



Task, T

- Imputasi nilai yang hilang:
 - Dalam jenis tugas ini, algoritma pembelajaran mesin diberi contoh baru $\mathbf{x} \in \mathbb{R}^n$, tetapi dengan beberapa entri x_i dari \mathbf{x} yang hilang.
 - Algoritma harus memberikan prediksi nilai entri yang hilang.
- Denoising:
 - Dalam jenis tugas ini, algoritma pembelajaran mesin diberikan sebagai input contoh rusak $\tilde{x} \in \mathbb{R}^n$ diperoleh dengan proses yang rusak yang tidak diketahui dari contoh bersih $x \in \mathbb{R}^n$.
 - Sistem harus memprediksi contoh bersih x dari versi rusaknya \tilde{x} , atau lebih umum memprediksi distribusi probabilitas bersyarat $p(x | \tilde{x})$



Task, T

- Estimasi kepadatan atau estimasi fungsi massa probabilitas:
 - Dalam masalah estimasi kepadatan, algoritma pembelajaran mesin diminta untuk mempelajari fungsi $p_{model}: \mathbb{R}^n \rightarrow \mathbb{R}$, di mana $p_{model}(x)$ dapat diartikan sebagai fungsi kepadatan probabilitas (jika x kontinu) atau fungsi massa probabilitas (jika x diskrit) pada ruang tempat contoh diambil.
 - Untuk melakukan tugas seperti itu dengan baik, algoritma perlu mempelajari struktur data yang telah dilihatnya. Itu harus tahu di mana contoh-contoh berkelompok dengan erat dan di mana mereka tidak mungkin terjadi.
 - Sebagian besar tugas yang dijelaskan di atas memerlukan algoritma pembelajaran untuk setidaknya secara implisit menangkap struktur distribusi probabilitas.



Performance Measure, P

- Untuk mengevaluasi kemampuan algoritma pembelajaran mesin, kita harus merancang ukuran kuantitatif kinerjanya.
- Biasanya ukuran kinerja **P** ini khusus untuk tugas **T** yang dilakukan oleh sistem.
- Untuk tugas seperti klasifikasi, klasifikasi dengan input yang hilang, dan transkripsi, kita sering mengukur **keakuratan (accuracy) model**.
- Akurasi hanyalah proporsi contoh yang modelnya menghasilkan keluaran yang benar.



Performance Measure, P

- Kita juga dapat memperoleh informasi yang setara dengan mengukur **tingkat kesalahan (error rate)**, proporsi contoh yang modelnya menghasilkan **output yang salah**.
- Kita sering menyebut tingkat kesalahan sebagai kerugian (**loss**) 0-1 yang diharapkan.
- Kerugian (loss) 0-1 pada contoh tertentu adalah 0 jika diklasifikasikan dengan benar dan 1 jika tidak.



Performance Measure, P

- Untuk tugas-tugas seperti estimasi kepadatan, tidak masuk akal untuk mengukur akurasi, tingkat kesalahan, atau jenis kerugian 0-1 lainnya.
- Sebagai gantinya, kita harus menggunakan **metrik performa** berbeda yang memberi model skor bernilai kontinu untuk setiap contoh.
- Pendekatan yang paling umum adalah melaporkan **probabilitas log rata-rata** yang diberikan model ke beberapa contoh.



Performance Measure, P

- Biasanya kita tertarik pada seberapa baik kinerja algoritma pembelajaran mesin pada data yang belum pernah dilihat sebelumnya, karena ini menentukan seberapa baik algoritma akan bekerja saat digunakan di dunia nyata.
- Oleh karena itu, kita mengevaluasi ukuran kinerja ini menggunakan kumpulan data pengujian (***test set***) yang terpisah dari data yang digunakan untuk melatih sistem pembelajaran mesin.



Experience, E

- Algoritma pembelajaran mesin dapat dikategorikan secara luas sebagai tidak diawasi (***unsupervised***) atau diawasi (***supervised***) oleh pengalaman (***experience***) seperti apa yang boleh mereka miliki selama proses pembelajaran.
- Sebagian besar algoritma pembelajaran dapat dipahami sebagai diizinkan untuk mengalami (***experience***) seluruh kumpulan data (data set).
- Terkadang disebut sebagai contoh titik data (***data points***).

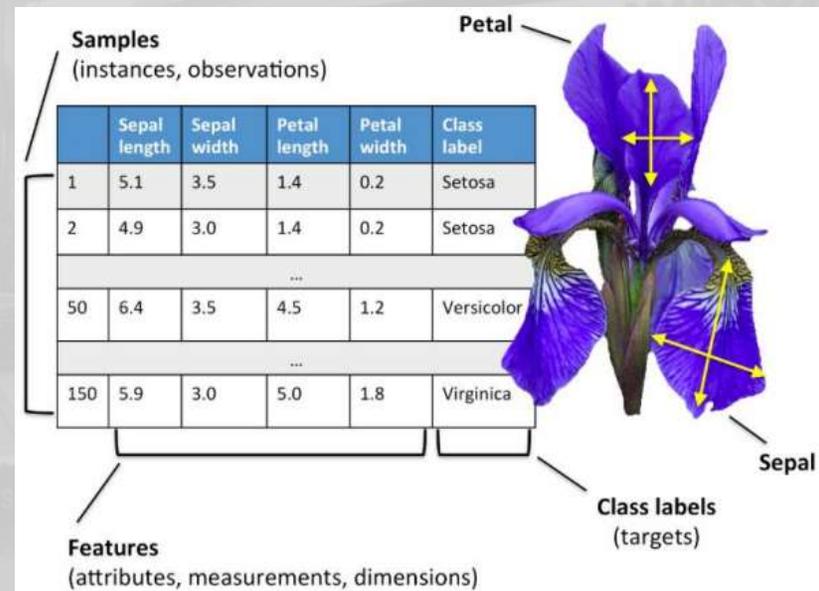


Experience, E

- Salah satu kumpulan data tertua yang dipelajari oleh ahli statistik dan peneliti pembelajaran mesin adalah kumpulan data Iris (Fisher, 1936).
- Ini adalah kumpulan pengukuran berbagai bagian dari 150 tanaman iris.
- Setiap tanaman individu sesuai dengan satu contoh.
- Ciri-ciri dalam setiap contoh adalah ukuran setiap bagian tumbuhan: panjang sepal, lebar sepal, panjang petal, dan lebar petal.
- Kumpulan data juga mencatat spesies mana yang dimiliki oleh setiap tanaman.
- Tiga spesies berbeda terwakili dalam kumpulan data.



Experience, E



Experience, E

- Algoritma pembelajaran tanpa pengawasan (***unsupervised***) mengalami (*experience*) kumpulan data yang berisi banyak fitur, lalu mempelajari properti berguna dari struktur kumpulan data ini.
- Dalam konteks pembelajaran mendalam, kita biasanya ingin mempelajari seluruh distribusi probabilitas yang menghasilkan kumpulan data, baik secara eksplisit, seperti dalam estimasi kepadatan, atau secara implisit, untuk tugas seperti sintesis atau denoising.
- Beberapa algoritma pembelajaran tanpa pengawasan lainnya melakukan peran lain, seperti pengelompokan, yang terdiri dari membagi kumpulan data menjadi kelompok contoh serupa.



Experience, E

- Algoritme pembelajaran yang diawasi (***supervised***) mengalami (*experience*) kumpulan data yang berisi fitur, tetapi setiap contoh juga dikaitkan dengan **label** atau **target**.
- Misalnya, kumpulan data Iris dianotasi dengan spesies dari setiap tanaman iris.
- Algoritma pembelajaran yang diawasi dapat mempelajari dataset Iris dan belajar mengklasifikasikan tanaman iris menjadi tiga spesies berbeda berdasarkan pengukurannya.



Experience, E

- Secara umum,
 - pembelajaran tanpa pengawasan melibatkan pengamatan beberapa contoh vektor acak x dan berusaha untuk secara implisit atau eksplisit mempelajari distribusi probabilitas $p(x)$, atau beberapa sifat menarik dari distribusi tersebut;
 - pembelajaran dengan pengawasan melibatkan pengamatan beberapa contoh vektor acak x dan nilai terkait atau vektor y , kemudian belajar memprediksi y dari x , biasanya dengan memperkirakan $p(y | x)$.



Experience, E

- Istilah:
 - pembelajaran terbimbing berasal dari pandangan target y yang diberikan oleh seorang instruktur atau guru yang menunjukkan sistem pembelajaran mesin apa yang harus dilakukan.
 - Dalam pembelajaran tanpa pengawasan, tidak ada instruktur atau guru, dan algoritma harus belajar memahami data tanpa panduan ini.



Experience, E

- Beberapa algoritma pembelajaran mesin tidak hanya mengalami (*experience*) kumpulan data tetap (*fixed*).
- Misalnya, algoritma pembelajaran penguatan (***reinforcement***) berinteraksi dengan lingkungan, sehingga ada umpan balik antara sistem pembelajaran dan pengalaman (*experience*) nya





TERIMA
KASIH



Universitas 17 Agustus 1945 Surabaya

Teknik Informatika