

Jawaban Implementasi Algoritma Supervised

Soal

- Apa itu *supervised learning*?
- Jelaskan bagaimana cara kerja dari algoritma yang anda implementasikan!
- Bandingkan ketiga algoritma tersebut, kemudian tuliskan kelebihan dan kelemahannya!
- Jelaskan penerapan dari algoritma *supervised*! (misal di bidang kesehatan atau industri)

Jawaban

Jawaban a

Supervised learning adalah metode belajar dalam pembelajaran mesin berdasarkan data yang disediakan dengan label. Label di sini maksudnya adalah *output*.

Jawaban b

K-Tetangga Terdekat (*K-Nearest Neighbor*)

Yang diperlukan adalah data uji, kueri yang ingin diprediksi nilainya, dan nilai K. Seperti namanya, cari K tetangga terdekat dengan menggunakan jarak Euclidean (tidak perlu akar kuadrat karena tidak memengaruhi kemonotonan), dan kembalikan modulusnya (bisa saja lebih dari satu jika datanya sedikit).

Regresi Logistik (*Logistic Regression*)

Ambil persamaan linear dengan banyak variabel sesuai dimensi inputnya (n), misalnya $b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$. Normalisasi datanya sehingga setiap input dan output berada pada rentang $[0, 1]$ (kalau tidak, terjadi *overflow error* saat menghitung sigmoid). Kemudian, coba prediksi *output* untuk setiap data, misalkan y_{pred} , dan bandingkan dengan hasil yang diharapkan, misalkan y . Misalkan inputnya adalah $\langle x_1, x_2, \dots, x_n \rangle$. Maka, gunakan rumus

$$b'_i = b_i + \text{learning rate} * (y - y_{pred}) * y_{pred} * (1 - y_{pred}) * x_i.$$

untuk i bilangan asli hingga n . Untuk $i = 0$, rumusnya sama dengan $x_0 = 1$. Lakukan sebanyak beberapa kali (*epoch*), kemudian gunakan persamaan linear tersebut untuk memprediksi kueri. Saya juga mencetak akurasi untuk menilai keandalan.

ID3 (*Iterative Dichotomiser 3*)

Intinya adalah mencari kolom yang membuat *data set* dipisah secara horizontal (berdasarkan nilai dari kolom tersebut) yang mengurangi ketidakpastian dalam output. Itu disebut entropi dan dihitung dengan rumus

$$\text{entropy}(\text{dataset}) = \sum_i -p_i \log_2(p_i)$$

dengan i adalah *output* yang ada di *data set*, p_i adalah proporsi antara banyak baris dengan output bernilai i dari seluruh baris. Pengurangan entropi digunakan untuk *information gain*. Caranya adalah hitung entropi data set sekarang dan pilih atribut yang jika dipisahkan secara horizontal, pengurangan antara entropi sebelumnya dengan rata-rata entropi hasil pemisahan menjadi maksimal. Nanti terbentuklah *decision tree*. Lakukan hal tersebut secara rekursif. Untuk implementasi saya, saya menggunakan prosedur saja dan *print* untuk membentuk *decision tree* berupa kondisi *if* yang banyak.

Jawaban c

- KNN mudah diimplementasikan, tetapi sulit digunakan karena perlu menentukan nilai K yang cocok. K yang rendah sangat peka dengan *noise*, sedangkan K yang tinggi sangat peka dengan pencilan. Selain itu, untuk kueri yang berbeda, perhitungannya juga berbeda. Cocok untuk data kontinu.
- *Logistic regression* tidak perlu mengulangi komputasi setiap ada kueri seperti KNN karena cukup menyimpan koefisien yang telah dikoreksi (tetapi di implementasi saya tidak simpan). Namun, yang saya lihat adalah ini tidak cocok untuk data yang tidak linear karena persamaannya saja linear. Berbeda dengan KNN yang perhitungannya berdasarkan tetangga di dekatnya. Cocok untuk data kontinu. Selain itu, *overfitting* juga berbahaya di sini karena koreksinya akan semakin lambat sehingga akurasi sulit untuk berubah.
- *ID3* dapat menghasilkan kesimpulan dari semua data berupa *decision tree* sehingga tidak perlu kueri apapun. Algoritme ini paling dapat diandalkan di antara yang lain. Namun, menurut saya, algoritme ini yang paling sulit diimplementasikan dibandingkan dengan yang lain. Selain itu, memori yang digunakan cukup banyak. Algoritme ini tidak cocok untuk data kontinu, tetapi cocok untuk data ordinal.

Jawaban d

Supervised learning dapat digunakan untuk mengambil keputusan dengan cara memprediksi berdasarkan data-data yang ada. Misalnya untuk kesehatan, orang yang diabetes atau *covid-19* dapat diidentifikasi dengan menggunakan data-data sebelumnya.