

PR 1 Pemelajaran Mesin

PR 1 terdiri dari Naive Bayes Classifier, Basic Statistic, dan Perceptron & Linear Regression.

A. Naive Bayes Classifier

1. Andi dan Budi berencana mengambil Mata Kuliah Pemelajaran Mesin (PM) semester depan. Mereka mendapat informasi bahwa kuliah PM cukup sulit. Oleh karena itu, mereka menggunakan model prediksi Naive Bayes untuk memprediksi apakah mereka akan lulus atau tidak. Jika diberikan dataset mahasiswa sebelumnya, bantulah Andi dan Budi untuk memprediksi apakah mereka akan lulus atau tidak.

Dataset Kelulusan MK Pemelajaran Mesin

Nama Mahasiswa	Prasyarat	Kehadiran	IPK	Status
Haryo	A	Malas	Dengan Pujian	Tidak Lulus
Sisi	C	Malas	Memuaskan	Tidak Lulus
Rezki	B	Malas	Dengan pujian	Lulus
Lisa	A	Rajin	Memuaskan	Lulus
Chandra	C	Rajin	Sangat Memuaskan	Lulus

(Rev 1) Untuk poin a-c, gunakan asumsi yang diambil oleh Naive Bayes Classifier dalam perhitungan peluang.

- a. Hitunglah $P(\text{Prasyarat}=A|\text{Status}=\text{Lulus})$!
- b. Hitunglah $P(\text{Pras}=B, \text{hadir}=\text{Rajin}, \text{IPK}=\text{Sgt Memuaskan}|\text{Status}=\text{Lulus})$!
- c. Hitunglah $P(\text{Status}=\text{tdk lulus}|\text{Pras}=C, \text{hadir}=\text{malas}, \text{IPK}=\text{memuaskan})$!
- d. Lakukan Prediksi terhadap data Andi dan Budi apakah akan lulus atau tidak!

Andi	B	Rajin	Dengan Pujian	?
Budi	A	Malas	Memuaskan	?

Karena jumlah data yang kecil ataupun memang karakteristik datanya, bisa saja terjadi probabilitas bernilai nol yang jika dikalikan dengan probabilitas-probabilitas lainnya tentu akan menghasilkan nilai nol.

- e. Apakah selama proses perhitungan dari a-d anda menemukan kasus nilai probabilitas 0 yang menyebabkan hasil kali dengan probabilitas lain menghasilkan 0 juga? Jika iya, bagaimanakah cara mengatasi hal tersebut?

f. Terapkan metode yang anda sebutkan pada bagian e untuk kasus yang terdapat probabilitas 0, bandingkan hasil sebelum dan sesudah diterapkan metode tersebut? Apa yang dapat anda simpulkan?

B. Basic Statistic

2. Buktikan bahwa jika x_i adalah variable acak independen, maka:

$$\text{var} \left(\sum_{i=1}^n x_i \right) = \sum_{i=1}^n \text{var}(x_i)$$

3. Hitunglah covariance matrix A berikut:

(Rev 1) Asumsikan matriks A adalah *data matrix*, dengan kata lain baris adalah observasi dan kolom adalah fitur.

$$A = \begin{bmatrix} 7 & 4 & 7 \\ 8 & 5 & 9 \\ 6 & 6 & 5 \end{bmatrix}$$

C. Perceptron & Linear Regression

4. Apa yang dimaksud dengan overfitting dan underfitting? bagaimana pengaruhnya terhadap model machine learning yang dihasilkan? serta bagaimana cara mengatasi keduanya?

5. Jelaskan kenapa *perceptron* dapat terperangkap atau 'stuck' di *local minima*? Bagaimana cara mengatasinya?

(Rev 2) Gambar tidak diminta.

6. Andi dan Budi ingin mencoba model prediksi yang lain agar lebih yakin dengan hasil prediksi yang ia peroleh sebelumnya. Bantulah mereka untuk membangun sebuah model prediksi dengan menggunakan *perceptron* dengan dataset yang sama pada no. 1. Karena data bersifat kategorikal, gunakan nilai-nilai dalam kurung berikut untuk mewakili setiap level pada fitur.

- Prasyarat : A (1), B (0.5), C(0)
- Kehadiran: Malas (0), Rajin (1)
- IPK : Dengan Pujian (1), Sangat Memuaskan (0.5), Memuaskan (0)
- Status: Lulus(1), tidak lulus (0)

Selain itu, diberikan inisialisasi parameter sebagai berikut:

- learning rate: 0.2
- initial weights/bobot = $[w_1, w_2, w_3] = [0.2, 0, 0]$
- Fungsi aktivasi, :
 $T = 0; 1 \text{ if } > T; 0 \text{ if } \leq T$

a. Lakukan *training* hingga nilai error yang diperoleh = 0.

- b. Untuk setiap iterasi, tampilkan nilai bobot yang baru, hasil prediksi setiap data, dan error yang diperoleh! Anda boleh menggunakan bantuan Ms.Excel untuk memudahkan perhitungan.

(Rev 2) Iterasi di sini adalah tiap datapoint, tuliskan bobot yang baru dan nilai lain yang diminta tiap pemberian datapoint ke model.

- c. Lakukan prediksi terhadap data Andi dan Budi (lihat no.1) berdasarkan model yang telah anda bangun dengan menggunakan perceptron!

7. *Linear Regression* adalah suatu metode regresi dengan mengambil fungsi berupa garis, *plane*, atau *hyperplane*, tergantung pada dimensi data, dengan jarak terdekat pada *datapoint*. Salah satu cara mendapatkan fungsi *linear regression* adalah dengan menggunakan operasi matriks secara langsung, atau biasa disebut *closed form solution*. Namun, pendekatan lain penyelesaian juga dapat dilakukan, yakni menggunakan prinsip *Gradient Descent*.

- a. Jelaskan kelebihan dan kekurangan solusi *closed form* sehingga diperlukan pendekatan numerikal.
- b. Lakukan penyelesaian *linear regression* untuk y sebagai fungsi dari x_1 dan x_2 , untuk dataset berikut ini hingga iterasi (epoch) kedua.

(Rev 2) Iterasi yang dimaksud adalah epoch (banyak kali training data diberikan ke model).

x1	x2	y
1	3	5
5	2	4
8	4	7

8. Diberikan dataset sebagai berikut :

X1	X2	Kelas
5.1	3.8	1
4.6	3.2	1
5.3	3.7	1
5.6	3.0	2
6.2	2.2	2
6.7	3.0	2

- a. Pecahkan masalah klasifikasi diatas menggunakan feedforward neural network dengan satu hidden layer dengan inisialisasi acak. Lakukan pembelajaran

sampai 2 iterasi (epoch) ! Gambarkan juga arsitektur yang anda rancang untuk menyelesaikan permasalahan diatas!

(Rev 2) Iterasi dikurangi menjadi 2. Iterasi yang dimaksud di sini adalah epoch (banyak kali training data diberikan ke model).

(Rev 3) Inisialisasi acak.

- b. Lakukan uji untuk data baru $X_1 = 4.4$, $X_2 = 3.2$