

TUGAS PENDAHULUAN PRAKTIKUM MODUL III
PEMBELAJARAN MESIN



Oleh

Nama : Muhammad Tijan Abdurrahman Nama : Muhammad Reza Putra Anisi

NIM :105220023 NIM :105220053

PROGRAM STUDI ILMU KOMPUTER
FAKULTAS SAINS DAN KOMPUTER
UNIVERSITAS PERTAMINA
2022

Pertanyaan :

1. Jelaskan apa yang Anda ketahui mengenai Linear Regresi dan Logistik Regresi!
Apa hubungan antara Linear Regresi dan Logistik Regresi tersebut?
2. Apa yang Anda ketahui mengenai *correlation* dan *covariance* serta hubungan antara keduanya? Lengkapi juga dengan persamaan matematikanya!
3. Apa yang Anda ketahui mengenai gradient descent? Lengkapi juga dengan penjelasan persamaan matematikanya!
4. Jelaskan arti fisis serta penggunaan dari metrik pengukuran:
 - i. akurasi,
 - ii. recall,
 - iii. sensitivity,
 - iv. precision,
 - v. specificity,
 - vi. false positive rate, dan
 - vii. false negative rate!

BONUS!!!

Sebutkan syarat-syarat untuk melakukan Regresi Linear dan Regresi Logistik! (masing-masing minimal 3)

Jawaban :

1. Linear Regression adalah metode statistik untuk memodelkan hubungan antara variabel dependen (Y) dan satu atau lebih variabel independen (X) yang bersifat kontinu. Tujuannya adalah untuk menemukan garis atau persamaan linear terbaik yang menggambarkan hubungan antara variabel independen dan dependen. Linear Regression digunakan untuk membuat prediksi numerik.

Logistic Regression adalah metode statistik yang digunakan untuk memodelkan hubungan antara variabel dependen yang bersifat biner atau kategorikal (Y) dan satu atau lebih variabel independen (X) yang bersifat kontinu atau kategorikal. Tujuannya adalah untuk menemukan fungsi logistik yang menggambarkan hubungan antara variabel independen dan probabilitas variabel dependen. Logistic Regression digunakan untuk membuat prediksi kategorikal.

Hubungan antara Linear Regression dan Logistic Regression adalah bahwa kedua metode menggunakan variabel independen untuk memprediksi variabel dependen. Namun, Logistic Regression digunakan untuk masalah klasifikasi, sedangkan Linear Regression digunakan untuk masalah regresi. Selain itu, fungsi logistik yang digunakan dalam Logistic Regression adalah fungsi logistik sigmoid yang dapat diubah menjadi bentuk linear.

2. Covariance adalah pengukuran seberapa erat dua variabel berhubungan secara linier. Covariance adalah angka positif jika kedua variabel bergerak dalam arah yang sama, angka negatif jika kedua variabel bergerak dalam arah yang berlawanan, dan nol jika tidak ada hubungan linier antara kedua variabel.

Persamaan matematika untuk covariance antara dua variabel X dan Y dengan n observasi adalah:

$$\text{Cov}(X, Y) = \sum[(X_i - \bar{X})(Y_i - \bar{Y})] / (n-1)$$

Correlation adalah ukuran seberapa kuat hubungan linier antara dua variabel. Correlation dapat berkisar dari -1 hingga 1. Correlation yang positif menunjukkan hubungan linier positif antara dua variabel, correlation yang negatif menunjukkan hubungan linier negatif antara dua variabel.

Persamaan matematika untuk correlation antara dua variabel X dan Y dengan n observasi adalah:

$$r(X, Y) = \text{Cov}(X, Y) / (\text{SD}(X) * \text{SD}(Y))$$

Hubungan antara covariance dan correlation adalah bahwa correlation adalah covariance yang dinormalisasi oleh standar deviasi masing-masing variabel.

3. Gradient descent adalah algoritma optimasi yang digunakan untuk menemukan nilai minimum dari sebuah fungsi dengan mengiterasi langkah demi langkah untuk mengoptimalkan parameter-model yang digunakan dalam machine learning.

Prinsip kerja gradient descent adalah dengan menghitung gradien (turunan) dari fungsi yang ingin dioptimalkan terhadap parameter yang ada. Kemudian, nilai gradien tersebut digunakan untuk menentukan arah dan besar langkah yang harus diambil untuk mencapai nilai minimum fungsi tersebut.

Persamaan matematika untuk menghitung langkah gradient descent adalah:

$$\theta = \theta - \alpha * \nabla J(\theta)$$

di mana θ adalah parameter-model yang ingin dioptimalkan, α adalah learning rate (tingkat pembelajaran), $\nabla J(\theta)$ adalah gradien dari fungsi biaya J terhadap parameter θ .

4. i. Akurasi

Akurasi mengukur seberapa banyak data yang diklasifikasikan dengan benar oleh model dibandingkan dengan total jumlah data yang ada. Artinya, akurasi adalah rasio antara jumlah prediksi yang benar dengan total jumlah prediksi yang dilakukan.

ii. Recall

Recall mengukur seberapa banyak data positif yang berhasil diklasifikasikan oleh model. Artinya, recall mengukur jumlah data positif yang diidentifikasi secara benar dibandingkan dengan total jumlah data positif yang sebenarnya ada.

iii. Sensitivity

Sensitivity, juga dikenal sebagai recall, mengukur seberapa banyak data positif yang berhasil diklasifikasikan oleh model. Artinya, sensitivity mengukur jumlah data positif yang diidentifikasi secara benar dibandingkan dengan total jumlah data positif yang sebenarnya ada.

iv. Precision

Precision mengukur seberapa akurat model dalam mengklasifikasikan data sebagai positif. Artinya, precision mengukur jumlah data positif yang diidentifikasi secara benar dibandingkan dengan total jumlah data yang diklasifikasikan sebagai positif oleh model.

v. Specificity

Specificity mengukur seberapa baik model dalam mengidentifikasi data negatif. Artinya, specificity mengukur jumlah data negatif yang diidentifikasi secara benar dibandingkan dengan total jumlah data negatif yang sebenarnya ada.

vi. False Positive Rate

False Positive Rate (FPR) mengukur seberapa sering model mengidentifikasi data negatif sebagai positif. Artinya, FPR mengukur jumlah data negatif yang salah diklasifikasikan sebagai positif dibandingkan dengan total jumlah data negatif yang sebenarnya ada.

vii. False Negative Rate

False Negative Rate (FNR) mengukur seberapa sering model mengidentifikasi data positif sebagai negatif. Artinya, FNR mengukur jumlah data positif yang salah diklasifikasikan sebagai negatif dibandingkan dengan total jumlah data positif yang sebenarnya ada.

Soal Bonus

Syarat-syarat untuk melakukan Regresi Linear:

Hubungan linear antara variabel independen dan dependen: Terdapat hubungan linear antara variabel independen dan dependen. Hal ini dapat dilihat dari scatter plot, di mana titik-titik data cenderung bergerak dalam pola garis lurus.

Homoskedastisitas: Varians residual atau error sama di semua titik data. Artinya, perbedaan antara nilai prediksi dan nilai aktual tidak bervariasi tergantung pada nilai variabel independen.

Normalitas: Data harus terdistribusi secara normal, yaitu bentuk distribusinya menyerupai kurva normal.

Syarat-syarat untuk melakukan Regresi Logistik:

Ketergantungan antara variabel dependen dan independen: Terdapat hubungan logistik antara variabel dependen dan independen, di mana variabel dependen dapat dijelaskan oleh variabel independen dalam bentuk logistik.

Data tidak memiliki outliers: Data tidak memiliki nilai yang ekstrim yang dapat mempengaruhi analisis dan hasil prediksi.

Tidak terjadi multikolinieritas: Variabel independen harus saling independen satu sama lain. Artinya, tidak ada hubungan kuat antara variabel independen yang dapat mempengaruhi hasil prediksi.