

# Logistic Regression

## 1. Cara kerja algoritma Logistic Regression

Algoritma Logistic Regression adalah algoritma klasifikasi yang digunakan untuk memprediksi probabilitas dari kelas biner (0 atau 1). Logistic Regression bekerja dengan menggunakan fungsi sigmoid untuk mengkonversi output linear dari model menjadi probabilitas antara 0 dan 1. Lebih runutnya berikut langkah-langkah utama dalam algoritma Logistic Regression.

### Inisialisasi:

- Inisialisasi bobot  $W$  dengan nol.
- Inisialisasi bias  $b$  dengan nol.
- Tentukan  $learning\_rate$ ,  $n\_iterations$ ,  $regularization$ ,  $C$ , dan  $loss\_function$ .

### Loop untuk Jumlah Iterasi:

- Hitung skor linear:  $z = X \cdot W + bz$
- Hitung prediksi probabilitas:  $y\_pred = 1/(1+e^{(-z)})$
- Hitung gradien:
  - Gradien bobot:  $dw = (1/m) \cdot X^T \cdot (y\_pred - y)$
  - Gradien bias:  $db = (1/m) \cdot \sum (y\_pred - y)$
  - Jika regularisasi l2, tambahkan penalti:  $dw += (C/m) \cdot W$
  - Jika regularisasi l1, tambahkan penalti:  $dw += (C/m) \cdot \text{sign}(W)$
- Perbarui parameter:
  - $W = W - learning\_rate \times dw$
  - $b = b - learning\_rate \times db$
- Setiap 100 iterasi, hitung dan cetak cost dengan fungsi loss yang dipilih.

### Prediksi:

- Untuk prediksi baru, hitung skor linear dengan parameter yang telah diupdate.
- Terapkan fungsi sigmoid untuk mendapatkan probabilitas.
- Klasifikasi ke kelas 0 atau 1 berdasarkan probabilitas dengan threshold 0.5.

Model melakukan iterasi sesuai jumlah yang ditentukan ( $n\_iterations$ ). Setiap iterasi, parameter diperbarui untuk meminimalkan fungsi loss, dan model akan terus belajar hingga kesalahan berkurang.

#### 4. Hasil Evaluasi Model

From Scratch	From Scikit-Learn
<pre>Accuracy: 0.8219461697722568       precision    recall  f1-score   support       0       0.84      0.97      0.90       399      1       0.45      0.11      0.17        84   accuracy          0.82       483  macro avg         0.64       483  weighted avg      0.77       483</pre>	<pre>Accuracy: 0.8322981366459627       precision    recall  f1-score   support       0       0.85      0.97      0.91       399      1       0.56      0.18      0.27        84   accuracy          0.83       483  macro avg         0.70       483  weighted avg      0.80       483</pre>

Implementasi scikit-learn sedikit lebih akurat dibandingkan implementasi dari scratch. Hasil evaluasi dari scikit-learn juga menunjukkan precision, recall, dan F1-score yang sedikit lebih baik untuk kedua kelas. Hal ini mengindikasikan bahwa implementasi scikit-learn lebih efektif dalam mengoptimalkan model dan menangani ketidakseimbangan kelas.

Implementasi dari scratch mungkin kurang stabil secara numerik dibandingkan dengan scikit-learn. Perhitungan yang tidak akurat dalam gradient descent, terutama untuk kasus di mana fitur sangat bervariasi, dapat mempengaruhi hasil model. Implementasi dari scikit-learn mungkin memiliki penanganan regularisasi yang lebih baik dan dapat melakukan tuning parameter secara otomatis untuk memberikan hasil yang lebih optimal. Selain itu, scikit-learn biasanya memanfaatkan pengaturan default yang telah teruji dan disesuaikan untuk berbagai dataset, sedangkan implementasi dari scratch mungkin memerlukan tuning parameter lebih lanjut untuk hasil yang optimal.

#### 5. Improvement

Untuk mengoptimalkan model dapat dilakukan pengaturan hyperparameter dan perbaikan algoritma serta implementasi. Hyperparameter tuning dapat dilakukan dengan menggunakan teknik seperti Grid Search atau Random Search untuk menemukan kombinasi optimal dari parameter model, termasuk nilai regularisasi C dan learning rate. Selain itu, eksperimen dengan berbagai jenis regularisasi, seperti L1 atau L2, dapat membantu mengatasi overfitting dan meningkatkan kemampuan generalisasi model. Penerapan learning rate scheduling, seperti pengurangan learning rate secara bertahap, juga dapat memperbaiki konvergensi model. Selain itu, memperbaiki stabilitas numerik dari fungsi sigmoid dan cost function dengan metode seperti regularisasi tambahan atau algoritma yang lebih stabil dapat mengurangi masalah numerik dan meningkatkan akurasi model.