

Seleksi Asisten Lab AI '23

Supervised Learning: Logistic Regression

1. Cara Kerja Algoritma Logistic Regression.

Konsep inti: Logistic Regression adalah model linear untuk klasifikasi yang memprediksi probabilitas sebuah instance termasuk dalam kelas tertentu. Prosesnya mengubah skor linear menjadi probabilitas menggunakan fungsi Sigmoid.

- **Model Linear:** Menghitung skor berbobot dari fitur input.

$$z = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n$$

- **Fungsi Sigmoid:** Mengubah skor z menjadi probabilitas antara 0 dan 1.

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Proses pelatihan untuk menemukan bobot terbaik menggunakan Gradient Descent.

Pseudocode untuk Prediksi KNN:

Algorithm 1

```

1: function TRAIN( $X, y, \alpha, n_{iters}$ )
2:   Inisialisasi bobot  $w \leftarrow$  vektor nol
3:   for  $i$  from 1 to  $n_{iters}$  do
4:      $z \leftarrow X \cdot w$ 
5:      $h \leftarrow \sigma(z)$ 
6:
7:      $gradient \leftarrow \frac{1}{m} \cdot X^T \cdot (h - y)$ 
8:
9:      $w \leftarrow w - \alpha \cdot gradient$ 
10:  end for
11:  return  $w$ 
12: end function

```

2. Analisis Perbandingan Eksperimen.

Table 1: Perbandingan kinerja

Metric	From Scratch	Scikit-learn
Accuration	0.8020	0.8041
Precision	0.65	0.65
Time (s)	0.15	0.03

Hasil performa kedua implementasi sangat mirip. Perbedaan minor terjadi karena scikit-learn menggunakan solver yang lebih canggih. Secara keseluruhan, model Logistic Regression (akurasi $\sim 80\%$) berkinerja lebih baik daripada model KNN (akurasi $\sim 76\%$) pada dataset ini.

3. *Improvement* yang bisa dilakukan untuk mencapai hasil yang lebih baik.

Beberapa bagian yang dapat ditingkatkan adalah

- *Regularization*
Menerapkan regularisasi L1 (Lasso) atau L2 (Ridge) untuk mencegah overfitting dengan cara memberikan "hukuman" pada bobot yang terlalu besar. Ini sangat berguna pada dataset dengan banyak fitur.
- *Imbalanced Dataset*
Sama seperti pada KNN, dataset ini tidak seimbang. Perlu digunakan teknik yang dapat membantu model untuk lebih baik dalam mengenali kelas minoritas (pelanggan yang *churn*), sehingga dapat meningkatkan nilai *recall*.