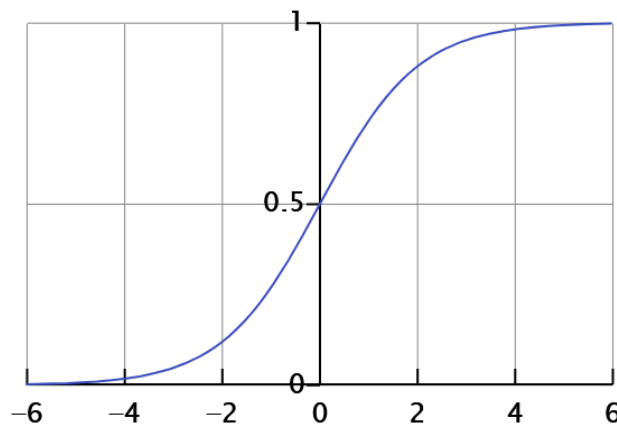


# Logistic Regression

Salman Hanif - 13523056

## Cara Kerja Algoritma

Algoritma Logistic Regression adalah linear regression yang dimodifikasi atau disesuaikan dengan fungsi sigmoid sedemikian agar keluaran berupa binary (0/1). Linear regression yang diubah dengan fungsi sigmoid akan berbentuk S-shaped curve, dan di situlah ditentukan rentang nilai yang menjadi 0 (0-0.5), dan rentang yang menjadi 1 (0.5-1).



Pertama, logistik regression membentuk prediksi linear menggunakan regresi linear. Bobot yang sudah dicari untuk setiap fitur dilakukan dengan fitur dan ditambahkan bias.

Nilai yang didapat sebelumnya kita masukkan ke fungsi sigmoid untuk menjadikannya di rentang 0-1.

Kemudian untuk optimasi model, dilakukan gradient descent yang dipandu Cross-Entropy Loss. intinya untuk bobot dari setiap fitur akan dikurangi dengan gradien dari fungsi cross entropy sehingga bobot dan bias bergerak berlawanan dan mengurangi loss dengan dikalikan besaran learning rate.

Fit di sini melakukan inisialisasi pada bobot dan bias yang mula-mula nilainya 0. Selama sebanyak jumlah iterasi, gradient descent diulang berkali kali untuk meningkatkan performa fitur.

Predict akan menggunakan bobot dan bias yang dipelajari dari fit, diubah sigmoid, kemudian ditentukan hasil prediksi.

### **Hasil Evaluasi model dari hasil scratch dan dari library. Jelaskan perbedaan**

Dalam evaluasi, digunakan metode Hold-out dan K-Fold pada library scikit-learn dan Logistic Regression buatan saya. dan performa model saya masih di bawah scikit learn.

Saat saya selidiki, logistic regression di scikit learn memiliki algoritma yang lebih kompleks, tidak hanya iterasi dengan learning rate dan gradient descent. Mungkin menggunakan implementasi Newton's method dalam optimasi karena parameter learning rate tidak ada saat saya coba samakan.

### **Improvement yang bisa dilakukan**

Improvement yang bisa saya lakukan adalah mengembangkan algoritma optimasi Newton's method yang ternyata merupakan sebuah bonus dalam pembuatan. Saya juga perlu menyesuaikan iterasi dan learning rate yang dinamis dan menyesuaikan jumlah data yang dilatih karena ternyata satu parameter yang tetap tidak bekerja dengan baik di semua data.