

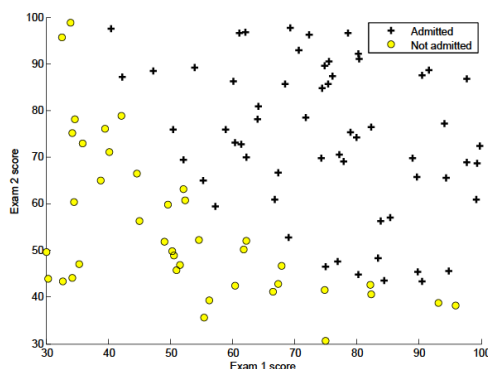
Assignment - Individual	
Course	Pengantar Pembelajaran Mesin (3 SKS)
Lecturer	Dr. Bambang Heru Iswanto
Submitted	(see Epsilon)
Instruction	<ul style="list-style-type: none">• Tulis nama anda dan NIM• Jawaban dan program dengan Jupyter Notebook/Google Colab• Upload jawaban ke Epsilon dalam format: ipynb dan .html

Session:

Klasifikasi dengan Logistic Regression

Misalkan Anda sebagai staf admisi diminta untuk menentukan besar peluang penerimaan mahasiswa baru berdasarkan hasil dua ujian. Anda diberi data historis *data2d40c2.txt* untuk melatih model klasifikator *regresi logistik*. Data sampel berisi pasangan input X (skor hasil dua ujian) dan y , yakni label kelas berupa keputusan penerimaan $\{1/0\}$. Tugas Anda adalah membangun model klasifikasi yang memperkirakan probabilitas penerimaan calon mahasiswa berdasarkan skor dari kedua ujian tersebut. Note: Sebelum memulai pelajari terlebih dahulu buku referensi. Kemudian lakukan langkah-langkah sbb.:

1. Lakukan plotting data untuk analisa awal data, tampilan misalnya seperti ini.



2. Gunakan fungsi sigmoid $g(z)$ sebagai fungsi hipotesis:

$$h_w(x) = g(z)$$

dimana

$$g(z) = \frac{1}{1+e^{-z}} \text{ dan } z = \theta^T x$$

Buatlah fungsi *sigmoid* tersebut dalam Python sehingga dapat dipanggil oleh fungsi lain dalam program Anda. Setelah selesai, cobalah uji beberapa nilai dengan memanggil `sigmoid(x)` pada baris perintah Python. Untuk nilai x positif yang besar, *sigmoid* harus mendekati 1, sedangkan untuk nilai negatif yang besar, *sigmoid* harus mendekati 0. Pastikan bahwa luaran *sigmoid*(0) = 0,5.

3. Sekarang tulis fungsi *costFun(theta, X, y)* pada Python untuk regresi logistik yang mengembalikan nilai *cost* dan *gradien*.
cost dihitung dengan persamaan:

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m [-y^{(i)} \log(h_{\theta}(x^{(i)})) - (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)}))]$$

gradien dihitung dengan persamaan:

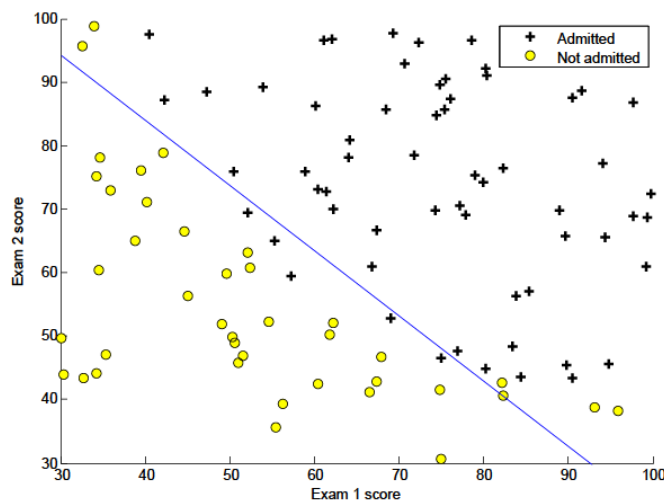
$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

4. Untuk membangun klasifikator regresi logistik, pertama inisialisasi nilai parameter θ dengan harga kecil sembarang, misalnya 0,1. Hitung berapa nilai *cost* dan *gradien* dengan data latih yang ada. Kemudian optimalkan fungsi $J(\theta)$ dengan update parameter θ dengan aturan *gradient ascent* :

$$\theta_j := \theta_j + \alpha (y^{(i)} - h_{\theta}(x^{(i)})) x_j^{(i)}$$

dimana α adalah *learning rate* yang bisa diset misalnya 0.2. Hitung kembali nilai *cost* dan *gradien*. Lakukan proses tersebut berulang hingga nilai *cost* optimal, atau perubahan sudah sangat kecil.

5. Setelah pelatihan selesai, coba gunakan model tersebut untuk memprediksi apakah seorang calon mahasiswa akan diterima jika skor ujian: 45 dan 85. Untuk itu Anda tulis fungsi *predik(X)* yang memberikan luaran nilai probabilitas penerimaan.
6. Buatlah plot data latih dan garis batas keputusan (*decision boundary*) pada satu grafik. Contoh tampilan seperti di bawah ini:



7. Ujilah model tersebut dengan test data: *datatest2d10c2.txt* . Buatlah matriks konfusi, kemudian hitunglah *accuracy*, *recall*, dan *F1-score*.

Referensi

1. Sebastian Raschka, "Modeling class probabilities via logistic regression", *Python Machine Learning 3rd Ed.*, Packt Pub., 2019, pp. 60-75.
2. Andrew Ng and Tengyu Ma, *CS229 Lecture Notes*, Stanford University, 2023.