Home Work #4: Logistic Regression

(Tugas dikerjakan sendiri dengan tulisan tangan dan dibuat softcopy pdf-nya yang masih mudah terbaca, tidak buram karena kurang cahaya atau terlihat kabur tulisannya. Pengumpulan dilakukan secara online di situs kuliah.itb.ac.id paling lambat hari Kamis, 1 Oktober pukul 23.55)

1. Diberikan $P(Y=0|X) = \frac{1}{1+\exp\,(w_0+\sum_{i=1}^{n} w_i X_i)}$ dan $P(Y=1|X) = \frac{\exp\,(w_0+\sum_{i=1}^{n} w_i X_i)}{1+\exp\,(w_0+\sum_{i=1}^{n} w_i X_i)}$

   (X^l, Y^l) adalah pasangan masukan dan keluaran dari data pelatihan ke-l.
   Jika $L(W) = \ln P(Y^l | X^l, W)$, tunjukkan bahwa $\frac{\partial L(W)}{\partial w_i} = \sum_i X_i^l (Y^l - \check{P}(Y^l|X^l, W))$.

2. Untuk model logistic regression multi-kelas, posterior probability diberikan oleh:

$$P(Y = y_k | X = x) = \frac{\exp\,(w_{k0}+\sum_{i=1}^{n} w_{ki} x_i)}{1+\sum_{m=l}^{K-1} \exp\,(w_{m0}+\sum_{i=1}^{n} w_{mi} x_i)} \quad \text{untuk } k = 1, \dots, K-1$$

$$P(Y = y_K | X = x) = \frac{1}{1+\sum_{m=l}^{K-1} \exp\,(w_{m0}+\sum_{i=1}^{n} w_{mi} x_i)}$$

   Dimana w_{mi} menyatakan bobot untuk kelas ke-m dengan $Y = y_k$ dan masukan x_i. Setiap input adalah sebuah vektor dengan dimansi n. Untuk mempermudah notasi, diasumsikan bahwa nilai w_{Ki} = 0 sehingga secara menyeluruh bisa kita tuliskan kedua persamaan di atas menjadi:

$$P(Y = y_k | X) = \frac{\exp\,(w_{k0}+\sum_{i=1}^{n} w_{ki} x_i)}{1+\sum_{m=l}^{K-1} \exp\,(w_{m0}+\sum_{i=1}^{n} w_{mi} x_i)} \quad \text{untuk } k = 1, \dots, K$$

   a. Berapa banyak parameter yang diperlukan dan apa parameter tersebut?
   b. Diberikan data pelatihan sebanyak S sampel: $\{(x^1, y^1), (x^2, y^2), \dots, (x^S, y^S)\}$, tuliskan secara eksplisit fungsi log likelihood dan sederhanakan semaksimal mungkin.

$$L(w_1, \dots, w_{K-1}) = \sum_{i=1}^{S} \ln P(Y = y_i, | X = x_i)$$

   c. Hitung dan sederhanakan gradient L untuk setiap w_k.
   d. Jika ditambahkan term regularisasi $\frac{\lambda}{2} \sum_{m=1}^{K-1} \sum_{i=1}^{n} w_{mi}^2$ sehingga fungsi obyektifnya menjadi :

$$F(w_1, \dots, w_{K-1}) = L(w_1, \dots, w_{K-1}) - \frac{\lambda}{2} \sum_{m=1}^{K-1} \sum_{i=1}^{n} w_{mi}^2$$

   Hitung gradient *F* untuk setiap w_k .