Formulas from Logistic Regression Document

Discriminative Linear Models – Logistic Regression

$$l(x) = \log \frac{f_{X|C}(x|h_1)}{f_{X|C}(x|h_0)} = w^T x + c$$

$$\log \frac{P(C = h_1|x)}{P(C = h_0|x)} = \log \frac{f_{X|C}(x|h_1)}{f_{X|C}(x|h_0)} + \log \frac{\pi}{1 - \pi} = w^T x + b$$

Logistic Regression Model

$$P(C = h_1|x, w, b) = e^{(w^T x + b)} P(C = h_0|x, w, b)$$

$$P(C = h_1|x, w, b) = \frac{e^{(w^T x + b)}}{1 + e^{(w^T x + b)}} = \frac{1}{1 + e^{-(w^T x + b)}} = \sigma(w^T x + b)$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ is the sigmoid function.

Sigmoid Function Properties

$$1 - \sigma(x) = \sigma(-x)$$

$$\frac{d\sigma(x)}{dx} = \sigma(x)(1 - \sigma(x))$$

Likelihood Estimation

$$P(C_1 = c_1, \ldots, C_n = c_n|x_1, \ldots, x_n, w, b) = \prod_{i=1}^{n} P(C_i = c_i|x_i, w, b)$$

$$y_i = P(C_i = 1|x_i, w, b) = \sigma(w^T x_i + b)$$

$$P(C_i = 0|x_i, w, b) = 1 - y_i = \sigma(-w^T x_i - b)$$

Log-Likelihood Function

$$L(w, b) = \prod_{i=1}^{n} y_i^{c_i} (1 - y_i)^{(1-c_i)}$$

$$\ell(w, b) = \sum_{i=1}^{n} [c_i \log y_i + (1 - c_i) \log(1 - y_i)]$$

Maximum Likelihood Estimation

$$w^*, b^* = \arg\max_{w,b} \ell(w, b)$$

Minimizing the negative log-likelihood:

$$J(w, b) = -\ell(w, b) = \sum_{i=1}^{n} -[c_i \log y_i + (1 - c_i) \log(1 - y_i)]$$

Binary Cross-Entropy

$$H(c_i, y_i) = -[c_i \log y_i + (1 - c_i) \log(1 - y_i)]$$

For general distributions:

$$H(P, Q) = -E_P[\log Q(x)] = -\sum_{x \in S} P(x) \log Q(x)$$

Logistic Loss Function

$$H(c_i, y_i) = \log(1 + e^{-z_i(w^T x_i + b)})$$

where $z_i = 2c_i - 1$.

Regularized Logistic Regression

$$\tilde{R}(w, b) = \frac{\lambda}{2}\|w\|^2 + \sum_{i=1}^{n} \log(1 + e^{-z_i(w^T x_i + b)})$$

$$R(w, b) = \frac{\lambda}{2}\|w\|^2 + \frac{1}{n}\sum_{i=1}^{n} \log(1 + e^{-z_i(w^T x_i + b)})$$

Multiclass Logistic Regression (Softmax)

$$P(C = k|x) = \frac{e^{w_k^T x + b_k}}{\sum_{j=1}^{K} e^{w_j^T x + b_j}}$$

Log-likelihood:

$$\ell(W, b) = \sum_{i=1}^{n} \log P(C_i = c_i | X_i = x_i, W, b)$$

Cross-Entropy for Multiclass:

$$H(z_i, y_i) = -\sum_{k=1}^{K} z_{ik} \log y_{ik}$$

Softmax Loss:

$$J(W, b) = -\sum_{i=1}^{n}\sum_{k=1}^{K} z_{ik} \log y_{ik}$$

Regularized Multiclass Logistic Regression:

$$R(W, b) = \Omega(W) + \frac{1}{n} J(W, b)$$

L2 Regularization:

$$\Omega(w_1, \ldots, w_N) = \frac{1}{2}\sum_{i} \|w_i\|^2$$