Multiclass logistic regression

November 9, 2020

- Probability of each label, per sample

$$\mathbf{Pr}(\hat{y}_i = k) = \frac{\exp(\theta_k^T x_i)}{\sum_{j=1}^{K} \exp(\theta_j^T x_i)}$$

- Likelihood given a label

$$\mathbf{Pr}(\hat{y}_i = k | y_i) = \begin{cases} \frac{\exp(\theta_k^T x_i)}{\sum_{j=1}^{K} \exp(\theta_j^T x_i)} & y_i = k \\ 0 & \text{else.} \end{cases}$$

- Log likelihood

$$\log \mathbf{Pr}(\hat{y}_i | y_i) = \theta_{y_i}^T x_i - \log \left(\sum_{j=1}^{K} \exp(\theta_j^T x_i) \right)$$

- Optimization problem

$$\underset{\theta_1,\ldots,\theta_K}{\text{maximize}} \; \frac{1}{m} \sum_{i=1}^{m} \log \mathbf{Pr}(\hat{y}_i | y_i) = \underset{\theta_1,\ldots,\theta_K}{\text{maximize}} \; \frac{1}{m} \sum_{i=1}^{m} \left(\theta_{y_i}^T x_i - \log \left(\sum_{j=1}^{K} \exp(\theta_j^T x_i) \right) \right)$$