Andy Hsu
204-423-373

ADMM for Class-Imbalanced Training of Binary Image Classifiers

Problem Formulation and ADMM Derivation

Dataset and Initial Problem Setup

The dataset used in this project is CIFAR-10, a collection of 32x32 color images from 10 different classes. The coloring model is RGB, bringing the total dimensionality of each image vector to $\boldsymbol{R}^{3072}$. The pixel values were normalized before training, and two of the image classes were selected for the binary classification. The basic training problem for the logistic regression classifier is as follows:

$$\min_{w} \sum_{i}^{N_{total}} \log\left(1 + \exp(-y_i w^T x_i)\right)$$

$w \in \boldsymbol{R}^{3072}$ represents the weights of the classifier, $x_i \in \boldsymbol{R}^{3072}$ represents the image pixel data for the ith training sample, and $y_i = \{-1,1\}$ represents the class label of the ith training sample.

ADMM Derivation with Convex Constraints

For convenience, the image vectors are concatenated into matrices:

$$X \in \boldsymbol{R}^{N_{total}\times 3072},\ X_{major} \in \boldsymbol{R}^{N_{major}\times 3072},\ X_{minor} \in \boldsymbol{R}^{N_{minor}\times 3072}$$

The constrained training approach adds two affine constraints. The first constraint is a change of variables to simplify the ADMM derivation. The second constraint states that the average classifier score $(y_i w^T x_i)$ of samples from the minority class must be greater than or equal to the average classifier score of samples from the majority class:

$$\min_{u,w,z}\ \delta_C(z) + \sum_{i}^{N_{total}} \log\left(1 + \exp(-y_i u_i)\right)$$

$$\textit{subject to}:$$
$$Xw - u = 0$$
$$m^T w - z = 0$$
$$\textit{where } m = -\frac{1}{N_{minor}} X_{minor}^T \mathbf{1} - \frac{1}{N_{major}} X_{major}^T \mathbf{1}$$

$$C = \{z \mid z \geq 0\}$$

Now the ADMM updates are derived using a as the dual variable for the first constraint and b as the dual variable for the second constraint. First, w is updated:

$$w^{(k+1)} = \underset{\widehat{w}}{argmin} \; ({a^{(k)}}^T X\widehat{w} + \frac{t}{2}\|X\widehat{w} - u^{(k)}\|_2^2 + \frac{t}{2}\|m^T\widehat{w} - z^{(k)}\|_2^2)$$
$$= \underset{\widehat{w}}{argmin} \; ({a^{(k)}}^T X\widehat{w} + \frac{t}{2}\left(\widehat{w}^T X^T X\widehat{w} - 2{u^{(k)}}^T X\widehat{w}\right) + \frac{t}{2}((m^T\widehat{w})^2 - 2z^{(k)}m^T\widehat{w})$$

This is a quadratic equation in $\widehat{w}$, therefore the update can be computed by setting the derivative to zero and finding a solution to a system of linear equations:

$$w^{(k+1)} = (tX^TX + (tmm^T))^+ q$$
$$where \; q = -X^Ta^{(k)} + tX^Tu^{(k)} + tz^{(k)}m$$

Next, the variables (u, z) are jointly updated. The augmented Lagrangian of u is separable:

$$u_i^{(k+1)} = \underset{\hat{u}_i}{argmin} \; (-a_i^{(k)}\hat{u}_i + \frac{t}{2}\hat{u}_i^2 - t\left(Xw^{(k+1)}\right)_i\hat{u}_i + \log(1 + \exp(-y_i\hat{u}_i)))$$
$$i = 1, \dots, N_{total}$$

Each of the functions in $\hat{u}_i$ are differentiable and can be minimized using a simple iterative procedure such as Newton's method.

Now the z update is derived:

$$z^{(k+1)} = \underset{\hat{z}}{argmin} \; \left(\delta_C(\hat{z}) - b^{(k)}\hat{z} + \frac{t}{2}\|\hat{z} - m^Tw^{(k+1)}\|_2^2\right)$$

$$= \underset{\hat{z}}{argmin} \; \left(\delta_C(\hat{z}) + \frac{t}{2}\left\|\hat{z} - m^Tw^{(k+1)} - \frac{1}{t}b^{(k)}\right\|_2^2\right)$$

$$= prox_{\frac{1}{t}h}\left(m^Tw^{(k+1)} + \frac{1}{t}b^{(k)}\right), where \; h(p) = \delta_C(p)$$
$$= \frac{1}{t}P_C(m^Tw^{(k+1)} + \frac{1}{t}b^{(k)})$$

The projection onto C is easy to compute as it is the projection onto the non-negative orthant. Finally, the dual variables are updated:

$$a^{(k+1)} = a^{(k)} + t(Xw^{(k+1)} - u^{(k+1)})$$
$$b^{(k+1)} = b^{(k)} + t(m^Tw^{(k+1)} - z^{(k+1)})$$