Homework 2 for #70240413
"Statistical Machine Learning"

Instructor: Prof. Jun Zhu

April 10, 2015

1 SVM

Consider the binary classification problem with training data $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^N$ ($\boldsymbol{x}_i \in \mathbb{R}^d$, $y_i \in \{0, 1\}$). Derive the dual problem of the following primal problem of linear SVM:

$$\begin{aligned} \min_{\boldsymbol{w}, b, \boldsymbol{\xi}} \frac{\lambda}{2}\|\boldsymbol{w}\|^2 \quad &+ \quad \sum_{i=1}^N \xi_i \\ s.t. \quad y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b) \quad &\geq \quad 1 - \xi_i \quad \forall\, i = 1, \ldots, N \\ \xi_i \quad &\geq \quad 0 \quad \forall\, i = 1, \ldots, N \end{aligned}$$

(Please note that we explicitly include the offset b here.)

2 IRLS for Logistic Regression

For a binary classification problem $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^N$ ($\boldsymbol{x}_i \in \mathbb{R}^d$, $y_i \in \{0, 1\}$), the probabilistic decision rule according to "logistic regression" is

$$P_{\boldsymbol{w}}(y|\boldsymbol{x}) = \frac{\exp(y\boldsymbol{w}^\top \boldsymbol{x})}{1 + \exp(\boldsymbol{w}^\top \boldsymbol{x})}. \tag{1}$$

And hence the log-likelihood is

$$\begin{aligned} \mathcal{L}(\boldsymbol{w}) \quad &= \quad \log \prod_{i=1}^N P_{\boldsymbol{w}}(y_i|\boldsymbol{x}_i) \quad (2) \\ &= \quad \sum_{i=1}^N \big(y_i \boldsymbol{w}^\top \boldsymbol{x}_i - \log(1 + \exp(\boldsymbol{w}^\top \boldsymbol{x}_i))\big) \quad (3) \end{aligned}$$

Please implement the IRLS algorithm to estimate the parameters of logistic regression

$$\max_{\boldsymbol{w}} \ \mathcal{L}(\boldsymbol{w}) \tag{4}$$

and the L2-norm regularized logistic regression

$$\max_{\boldsymbol{w}} \; -\frac{\lambda}{2}\|\boldsymbol{w}\|_2^2 + \mathcal{L}(\boldsymbol{w}), \tag{5}$$

where λ is the positive regularization constant.

You may refer to the lecture slides for derivation details but you are more encouraged to derive the iterative update equations yourself.

Please compare the results of the two models on the "UCI a9a" dataset[1]. The suggested performance metrics to investigate are e.g. prediction accuracies (both on training and test data), number of IRLS iterations, L2-norm of $\|\boldsymbol{w}\|_2$, etc.

You may need to test a range of λ values with e.g. cross validation for the regularized logistic regression. You may also need to do PCA dimension reduction of the input data to make sure that the Hessian matrix is invertible.

3 Exponential Family

For an exponential family distribution:

$$p(\boldsymbol{x}|\boldsymbol{\eta}) = h(\boldsymbol{x}) \exp\left(\boldsymbol{\eta}^\top T(\boldsymbol{x}) - A(\boldsymbol{\eta})\right)$$

1. Derive both the 1st and 2nd order derivatives of $A(\boldsymbol{\eta})$ in its general form and show how they relate to the mean and (co)-variance of $T(\boldsymbol{x})$.

2. Demonstrate your above results on the the following special cases: (a) Bernoulli distribution, (b) multivariate normal distribution and (c) Dirichlet distribution.

[1] `http://ml.cs.tsinghua.edu.cn/~wenbo/data/a9a.zip`