A Maximum K-Min Approach for Classification

Mingzhi Dong and Liang Yin
Beijing University of Posts and Telecommunications, China
mingzhidong@gmail.com, yin@bupt.edu.cn

Abstract
In this paper, a general Maximum K-Min approach for classification is proposed, which focuses on maximizing the gain obtained by the K worst-classified instances while ignoring the remaining ones. To make the original optimization problem with combinational constraints computationally tractable, the optimization techniques are adopted and a general compact representation lemma is summarized. Based on the lemma, a Nonlinear Maximum K-Min (NMKM) classifier is presented and the experiment results demonstrate the superior performance of the Maximum K-Min Approach.

Introduction
In the realm of classification, maximin approach, which pays strong attention to the worst situation, is widely adopted and it is regarded as one of the most elegant ideas. Hard-margin Support Vector Machine (SVM) (Vapnik 2000) is the most renowned maximin classifier and it enjoys the intuition of margin maximization.

However, maximin methods based on the worst instance may be sensitive to outliers/noisy points near the boundary, as shown in Figure 1(a). Therefore, after the previous work on a special case of naive linear classifier (Dong et al. 2012), in this paper, we propose a general Maximum K-Min approach for classification, which focuses on maximizing the gain obtained by the K worst-classified instances while ignoring the remaining ones, as exemplified in Figure 1(c)(d).

Maximum K-Min Approach
Maximum K-Min Gain can be readily solved via Minimum K-Max Loss. Since Minimum K-Max are more frequently discussed in the realm of optimization, in the following sections, the Maximum K-Min approach will be discussed and solved via minimizing K-Max Loss.

With the prediction function \( f(x_n, w) \) of a classifier, \( g_n = t_n f_n = t_n f(x_n, w) \) represents the classification confidence of the \( n \)th training instance \( x_n \); \( g_n \geq 0 \) indicates the correct classification of \( x_n \) and the larger \( g_n \) is, the more confident the classifier outputs the classification result. Therefore, we can define a vector \( g = (g_1, \ldots, g_N)^T \in \mathbb{R}^N \).

Figure 1: A comparison of Maxmin Approach (Hard-Margin SVM) and linear Maximum K-Min Approach(Support Vectors/Worst K Instances are marked ×).

Original Objective Function
Thus, the original objective function of Maximum K-Min approach is to minimize the K-Max Loss and has the following formulation

\[
\min \quad \Theta_K = \sum_{i=1}^{K} \theta_{[i]},
\]

Copyright © 2013, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.
Equation 1 is a convex optimization problem as follows:
\[
\begin{align*}
\min_{s, y} & \quad s, \\
\text{s.t.} & \quad y_{i_1} + \cdots + y_{i_k} \leq s, \\
& \quad 1 \leq i_1 \leq \cdots \leq i_k \leq N,
\end{align*}
\]

Compact Representation of $\Theta_K$

However, it’s prohibitive to solve an optimization problem under $C_N^K$ inequality constraints. We need to sketch out a compact formulation. Thus, with the convex optimization techniques (Boyd and Vandenberghe 2004), we introduce the following lemma.

**Lemma 1** For a fixed integer $K$, $\Theta_K(y)$, the sum of $K$ largest elements of a vector $y$, is equal to the optimal value of a linear programming problem as follows,
\[
\begin{align*}
\max_{y} & \quad y^T z, \\
\text{s.t.} & \quad 0 \leq z_i \leq 1, \\
& \quad \sum_i z_i = K,
\end{align*}
\]

where $i = 1, \ldots, N, z = \{z_1, \ldots, z_N\}$.

According to Lemma 1, we can obtain an equivalent representation of $\Theta_K$. The dual of this equal representation can be written as:
\[
\begin{align*}
\min_{s, u_i} & \quad Ks + \sum_i u_i \\
\text{s.t.} & \quad s + u_i \geq y_i, \\
& \quad u_i \geq 0
\end{align*}
\]

According to the above conclusion, the minimization of $\Theta_K(y)$, i.e. the original Maximum K-Min problem, is equal to the minimization of $Ks + \sum_i u_i$ under $2N$ constraints.

**A Nonlinear Maximum K-Min Classifier**

To utilize kernel in non-linear separable situations, we make a linear assumption of the prediction function $f(x)$
\[
f(x) = \sum_{n=1}^{N} a_n k(x, x_n)t_n + b, \quad n = 1, \ldots N \quad (2)
\]

where $k(x, x_n)$ denotes a similarity measurement between $x$ and $x_n$ via a kind of kernel function. $a = \{a_1, \ldots, a_N\}$ denotes the weight of training samples in the task of classification. $b$ indicates the bias. $t_n$ indicates the corresponding target value and $t_n \in \{-1, 1\}$. The prediction of $x$ is made by taking weighted linear combinations of the training set target values, where the weight is the product of $a_n$ and $s(x, x_n)$. Therefore, the classification confidence measurement $g(x)$ equals $t_n f(x)$.

According to the above discussion, we can obtain a compact representation for this Nonlinear Maximum K-Min Classifier (NMKM) as follows:
\[
\begin{align*}
\min_{a, s, u_i} & \quad Ks + \sum_{n=1}^{N} u_i, \\
\text{s.t.} & \quad \sum_{n=1}^{N} a_n t_n k(x_i, x_n) + t_n b - s - u_i \leq 0, \\
& \quad u_i \geq 0, \\
& \quad \alpha^T \alpha \leq 1, \\
& \quad i = 1, \ldots, N,
\end{align*}
\]

As a Quadratically Constrained Linear Programming (QCLP) problem, standard convex optimization methods, such as interior-point methods (Boyd and Vandenberghe 2004), can also be adopted to solve the above problem efficiently and a global maximum solution is guaranteed.

**Experiment**

The algorithm of NMKM (with RBF kernel) is compared with two state-of-the-art classifiers of Support Vector Machine (SVM) (with RBF kernel) and Logistic Regression (LR) (L2-regularized). Ten UCI datasets for binary classification are adopted in the experiment. Our algorithm is deployed using cvx toolbox\(^1\) in matlab environment and kernel SVM is deployed using libsvm toolbox\(^2\). The hyperparameters of NMKM and SVM are chosen via cross-validation.

As shown in Table 1, among 10 different datasets, NMKM achieves best accuracy in 6 datasets, SVM performs best in 2 datasets and LR shows best performance in 2 datasets. Among all datasets which SVM or LR performs better than NMKM, NMKM performs slightly weaker than SVM or LR. While for datasets which NMKM gains better performance, the accuracy of NMKM may be much larger than MSVM and LR. The accuracy gap of SVM in dataset liver and LR in dataset fourclass from NMKM is 8.86% and 27.85% respectively.

Thus we can conclude, compared with SVM and LR, NMKM has shown competitive classification performance, which verifies the efficiency of Maximum K-Min approach.

**References**


---

2. www.csie.ntu.edu.tw/~cjlin/libsvm