NON-DECOMPOSABLE PERFORMANCE MEASURES

Debojyoti Dey (15511264)
Nimisha Agarwal (15511267)
(Group 15)

November 15, 2016

Overview

Problem Statement

Finding general optimization techniques which can work on all non-decomposable performance measures.

Misclassification

- Can be represented as -

$$\sum \frac{1 - y_i y_i^*}{2}$$

- Decomposable.

Multivariate

$$\overline{h} : \overline{\mathcal{X}} \rightarrow \overline{\mathcal{Y}}$$

$$\overline{h}_w(\overline{x}) = \underset{\overline{y}^* \in \mathcal{Y}}{\arg\max}(w^T \psi(\overline{x}, \overline{y}^*))$$

$$\psi(\overline{x}, \overline{y}^*) = \sum_{i=1}^{n} y_i^* x_i$$

Non-decomposable Performance Measures

- Performance Measures which do not decompose linearly.
- In classification problems, error rates are generally used.
- But not useful in case of class imbalance.
- So, measures like -
  - F-measure is used for text retrieval.
  - A continuous function of TPR and TNR is used in class imbalanced classification settings.

Examples

G-mean:

$$\sqrt{pq}$$

Min

$$min(P, N)$$

H-mean:

$$\frac{2PN}{P+N}$$

Problems in Non-decomposable Performance Measures

- Non-convex.
- Require an approximate upper bound on the raw performance measure curve given by some (surrogate) convex function which can be optimized easily.

SVM Approach to Optimize Non-Linear Performance Measures

- **Training Sample**

$$\mathcal{S} = ((x_1, y_1), \cdots\cdots, (x_n, y_n)) \in (\mathcal{X} \times \mathcal{Y})^n$$

drawn i.i.d. according to some unknown probability distribution.

- Find a rule $h \in \mathcal{H}$ from hypothesis space $\mathcal{H}$ such that it optimizes the expected prediction performance

$$R^{\Delta}(h) = \int \Delta((h(x_1), \cdots, h(x_n)), (y_1, \cdots, y_n)) dPr(S)$$

SVM Approach to Optimize Non-Linear Performance Measures

- Loss function Δ is not decomposable.
- Optimize the empirical loss function Δ over the entire sample set.

$$\hat{R}_S^{\Delta}(h) = \Delta((h(x_1), \cdots, h(x_n)), (y_1, \cdots, y_n))$$

- For such problems, structural SVM is used.

SVM Approach to Optimize Non-Linear Performance Measures

General approach to this problem based on SVM

$$\min_{w,\zeta\geq 0} \quad \frac{1}{2}\|w\|^2 + C\xi$$

$$s.t \quad \forall \overline{y}^* \in \overline{\mathcal{Y}} \setminus \overline{y} : w^T[\psi(\overline{x},\overline{y}) - \psi(\overline{x},\overline{y}^*)] \geq \Delta(\overline{y}^*,\overline{y}) - \xi$$

$$\Rightarrow \Delta(\overline{y}^*,\overline{y}) + w^T(\Sigma x_i {y_i}^* - \Sigma x_i y_i) \leq \xi$$

$$\Rightarrow \Delta(\overline{y}^*,\overline{y}) + (\Sigma({y_i}^* - y_i) w^T x_i) \leq \xi$$

where $\Delta(\overline{y}^*,\overline{y})$ is loss function.

SVM Approach to Optimize Non-Linear Performance Measures

- Considering concave performance measures

$$P_\psi = \psi(TPR, TNR)$$

- For any concave function ψ and $\alpha, \beta \in \mathbb{R}$

$$\psi^*(\alpha, \beta) = \inf_{u,v \in \mathcal{R}} \{\alpha u + \beta v - \psi(u, v)\}$$

- By concavity,

$$\psi(u, v) = \inf_{\alpha,\beta \in \mathcal{R}} \{\alpha u + \beta v - \psi^*(\alpha, \beta)\}$$

SPADE Algorithm

- Requires link functions to be L-Lipschitz.
- Works with non-Lipschitz functions when some restrictions are imposed on them.

Our Solution

Trying to remove the restrictions to give a general solution.

$$\hat{R}_S^{\Delta}(h) = \Delta((h(x_1), \cdots, h(x_n)), (y_1, \cdots, y_n))$$

$$h(x, w) = \arg\max_{y \in \mathcal{Y}} F(x, y; w)$$

$$F(x, y; w) = \langle w, \psi(x, y) \rangle$$

$$h_w(\overline{x}, w) = \arg\max_{\overline{y} \in \mathcal{Y}} (w^T \psi(\overline{x}, \overline{y}))$$

$$\underset{w, \zeta \geq 0}{min} = \frac{1}{2} \|w\|^2 + C\zeta$$

$$s.t \quad \forall \overline{y}' \in \overline{\mathcal{Y}} \setminus \overline{y}^* : w^T [\psi(\overline{x}, \overline{y}) - \psi(\overline{x}, \overline{y}^*)] \geq \Delta(\overline{y}^*, \overline{y}) - \zeta$$

References

Joachims, Thorsten
A Support Vector Method for Multivariate Performance Measures.
Proceedings of the 22Nd International Conference on Machine Learning

Narasimhan, Harikrishna and Vaish, Rohit and Agarwal, Shivani
On the Statistical Consistency of Plug-in Classifiers for Non-decomposable Performance Measures.
Advances in Neural Information Processing Systems 27

Narasimhan, Harikrishna and Kar, Purushottam and Jain, Prateek
Optimizing Non-decomposable Performance Measures: A Tale of Two Classes.
ICML

Questions?