# NON-DECOMPOSABLE PERFORMANCE MEASURES

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

November 13, 2016

# Overview

# Problem Statement

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

# 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 which are used in class imbalanced classification settings.

# Problems in Non-decomposable Performance Measures

- Non-convex.
- Convex optimization techniques can not be applied on them.
- So, 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 $\Delta$ is not decomposable.
- Optimize the empirical loss function $\Delta$ 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, Structured SVM is used.

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

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

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

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

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

$$s.t \quad \forall \overline{y}' \in$$

$$\overline{\mathcal{Y}} \setminus \overline{y} : w^T[\psi(\overline{x}, \overline{y}) - \psi(\overline{x}, \overline{y}')] \geq \Delta(\overline{y}', \overline{y}) - \zeta$$

# 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?