Multi-Class and Multi-Label Classification

CS4780/5780 – Introduction to Machine Learning

Thorsten Joachims
Cornell University

Multi-Class Classification

- Binary Classification
  - $Y = \{+1, -1\}$
- Multi-Class Classification
  - $Y = \{1,2,3, \dots, k\}$
  - Example: Image classification as Cat vs. Dog vs. … vs. Tree
- Approaches
  1. "Native" multi-class methods: KNN, Naïve Bayes, LDA
  2. Multi-class extensions: multi-class SVMs, multi-class neural networks, …
  3. Multi-class to binary reductions
     - One-against-the-rest
     - Pairwise
     - Error Correcting Output Codes

One-Against-the-Rest

- Idea: Split multi-class problem into multiple binary classification problems.
- Procedure
  - Training:

    For each t ∈ $\{1,2,3,\ldots,k\}$, train binary classifier h_t with labels

$$\bar{y}_i = \begin{cases} +1 & if\ t = y_i \\ -1 & otherwise \end{cases}$$

  - Prediction:

$$h(\vec{x}) = \underset{t \in \{1,..,k\}}{\operatorname{argmax}} [\vec{w}_t \cdot \vec{x} + b_t]$$

Multi-Label Classification

- Multi-Class Classification
  - $Y = \{1,2,3, \dots, k\}$
  - Example: Image classification as one of Cat vs. Dog vs. Tree → $y = Cat$
- Multi-Label Classification
  - $Y = PowerSet\{1,2,3, \dots, k\}$
  - Example: Identify set of all objects in image $y = \{Cat, Dog\}$
- Procedure
  - Idea: Represent set of labels as bitvector $y = (1,1,-1)$
  - Training:

    For each t ∈ $\{1,2,3, \dots, k\}$, train binary classifier h_t with labels

$$\bar{y}_i = \begin{cases} +1 & if\ t \in y_i \\ -1 & otherwise \end{cases}$$

  - Prediction:

$$h(\vec{x}) = \left(h_1(\vec{x}), h_1(\vec{x}), \dots, h_k(\vec{x})\right)$$

Summary

- Multi-class and Multi-label problems can be reduced to binary classification
- Efficiency can be an issue, if the number of classes/labels is large
  (→ Extreme Classification)