Machine Learning 2018 – Multi-class Classification and Kernel Methods

Kien C Nguyen

December 5, 2018

Contents

1 Introduction

2 Multi-class Classification

3 References

Binary Classification

- Recall that in binary classification, we have to predict whether an element belongs to one of the two classes (Negative and Positive).
- Classify an email as Not Spam / Spam
- In credit scoring, classify a customer as Good / Bad
- In network intrusion detection, classify a connection as Normal / Attack
- Detect the gender (Male / Female) using profile pictures

- Learning
  The training set is $\mathbb{T} \triangleq \left(\left(\mathbf{x}^{(1)}, y^{(1)}\right), \left(\mathbf{x}^{(2)}, y^{(2)}\right), \ldots \left(\mathbf{x}^{(N)}, y^{(N)}\right)\right)$.
  where $\mathbf{x}^{(i)}$ is given by $\mathbf{x}^{(i)} = \begin{bmatrix} x_0^{(i)} \\ x_1^{(i)} \\ \ldots \\ x_D^{(i)} \end{bmatrix}$
  $x_0^{(i)} = 1$, $y^{(i)} \in \{0, 1, \ldots, k\}$
- Prediction
  Given a point $\mathbf{x}$, output a single class label y.

Multi-class Classification

- In this lecture, we look into methods to reduce a multi-class classification problem to multiple binary classification problems.
  - One-vs-All
  - All-vs-All
  - Error correcting codes

Contents

1 Introduction

2 Multi-class Classification

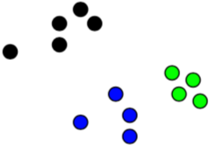
3 References

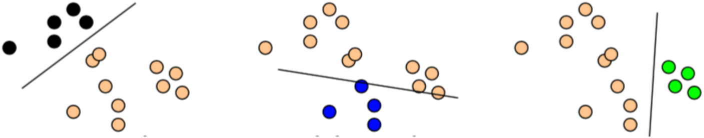
One-Vs-All

- Assumption: Each class can be separated from all the rest using a binary classifier
- Learning:
  - Decomposed to learning k independent binary classifiers, one for each class label.
  - For each label $l \in \{0, 1, \ldots, k\}$, construct a binary classification problem as follows
  - Positive examples: Elements of $\mathbb{T}$ with label l.
  - Negative examples: All other elements of $\mathbb{T}$
  - This is a binary classification problem that we can solve.
  - We will have k binary classifiers $\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \ldots, \mathbf{w}^{(k)}$.
- Decision: Winner Takes All (WTA):
  $l = argmax_i \left(\mathbf{w}^{(i)T}\mathbf{x}\right)$

One-Vs-All

- MultiClass classifier
  - Function $f : R^n \rightarrow \{1,2,3,\ldots,k\}$
- Decompose into binary problems





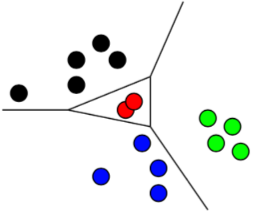
Source: UIUC – CS446

All-Vs-All

- Assumption: There is a separation between every pair of classes using a binary classifier
- Learning: Decomposed to learning $\binom{k}{2}$ independent binary classifiers, one corresponding to each pair of class labels. For the pair (i, j):
  - Positive example: all examples with label i
  - Negative examples: all examples with label j
- Decision: Each label gets $k - 1$ votes
- Decision Options:
  - Majority: classify example $\mathbf{x}$ to take label i if i wins on $\mathbf{x}$ more often than j $(j = 1, \ldots, k)$
  - A tournament: start with $n/2$ pairs; continue with winners

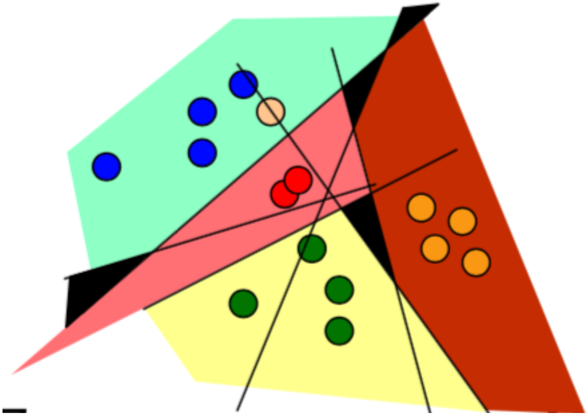
All-Vs-All

- Find $v_{rb}, v_{rg}, v_{ry}, v_{bg}, v_{by}, v_{gy} \in \mathbf{R}^d$ such that
  - $v_{rb}.x > 0$ if y = red
    < 0 if y = blue
  - $v_{rg}.x > 0$ if y = red
    < 0 if y = green
  - ... (for all pairs)



Source: UIUC – CS446

All-Vs-All Decision Regions



Source: UIUC – CS446

Error Correcting Codes

- Matrix of dimension $k \times n$ where $n > \lceil log_2 k \rceil$.
- Rows: An encoding of each class (k rows)
- Columns: L dichotomies (dichotomizers) of the data, each corresponds to a new classification problem
- Each training example is mapped to one example per column $(x, 3) \rightarrow \{(x, P1), +; (x, P2), -; (x, P3), -; (x, P4), +\}$
- To classify a new example x:
  - Evaluate hypothesis on the 4 binary problems $\{(x, P1), (x, P2), (x, P3), (x, P4)\}$
  - Choose label that is most consistent with the results. Use Hamming distance (bit-wise distance)

er column

Label	P1	P2	P3	P4
1	-	+	-	+
2	-	+	+	-
3	+	-	-	+
4	+	-	+	+
k	-	+	-	-

Source: UIUC – CS446

Contents

1. Introduction
2. Multi-class Classification
3. References

References

[1] Bishop, C. M. (2013). Pattern Recognition and Machine Learning.

[2] University of Illinois at Urbana-Champaign – CS446 Machine Learning