Multi-class classification

Reminder: linear SVM

Objective: maximize the margin while correctly classifying all examples correctly

$$\underset{\mathbf{w},b}{\text{minimize}} \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n}\xi_i$$

$$\text{subject to: } y_i(\mathbf{w}^\intercal \mathbf{x}_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0, \quad i = 1, \ldots, n\,.$$

Handling more than two classes

- SVM formulation designed for multi-class problems:

$$\text{minimize} \frac{1}{2}\sum_{i=1}^{k}||\mathbf{w}_i||^2 + C\sum_{i=1}^{n}\sum_{m\neq y_i}\xi_i^m$$

$$\text{subject to: } \mathbf{w}_{y_i}^{\mathsf{T}}\mathbf{x}_i + b_{y_i} \geq \mathbf{w}_m^{\mathsf{T}}\mathbf{x}_i + b_m + 1 - \xi_i^m, \ m \neq y_i,$$

$$\xi_i \geq 0, \quad i = 1, \ldots, n\,.$$

k – number of classes

To classify: $f(\mathbf{x}) = \underset{i}{\text{argmax}}\ \mathbf{w}_i^{\mathsf{T}}\mathbf{x} + b_i$

- There is a more efficient multi-class SVM formulation by Crammer and Singer (one slack variable per example).

Weston, Jason, and Chris Watkins. Multi-class support vector machines. Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, May, 1998.

Multi-class classification with binary classifiers

Alternative:

- Use a binary classifier to do multi-class classification.

How to construct a multi-class classifier out of binary classifiers?

There multiple ways of doing this!

one-versus-one

Approach:

- Train k(k-1)/2 classifiers on each pair of classes
- Testing: classify example to the class that receives the largest number of votes

In scikit-learn:

There is a multiclass module with a class that implements this strategy (multiclass.OneVsOneClassifier)

one-versus-the-rest

Approach:

- Train k binary classifiers, where classifier i is trained to separate class C_i from the rest
- In testing: assign a class label according to the classifier that provides the highest score to a test example

Simple and accurate!

See:

In Defense of One-Vs-All Classification. ***Ryan Rifkin, Aldebaro Klautau;*** Journal Of Machine Learning Research 5:101-141, 2004.

http://jmlr.org/papers/v5/rifkin04a.html

In scikit-learn: multiclass.OneVsRestClassifier

If you feed an svm instance multi-class data, it will automatically use this strategy.

Evaluating multi-class classification

How to evaluate multi-class classifiers?

Some measures of classifier accuracy such as area under the ROC curve are specific to binary classifiers.

scikit-learn demo