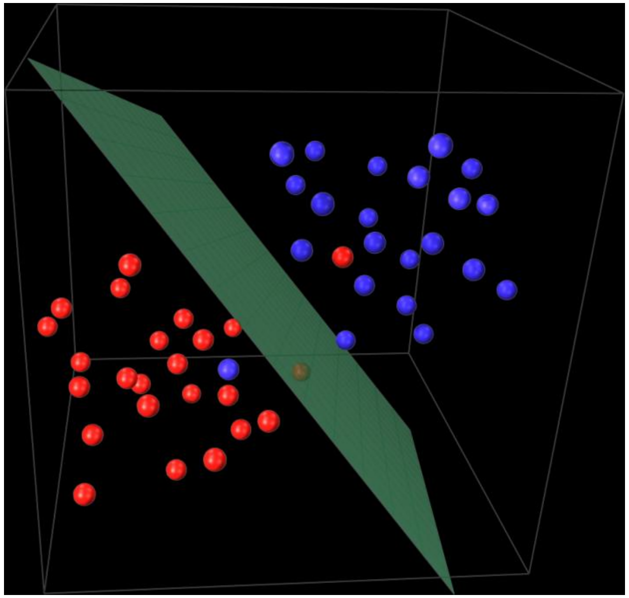
Multiclass SVM

Group # 2

MachineLearning@HackerDojo

Date: 11/11/2010

Support Vector Machine



Standard SVMs are typically designed only for binary classification (or 2 classes only)

Today we will look at the different methods to effectively use it for multi-class classification.

Multiclass methods

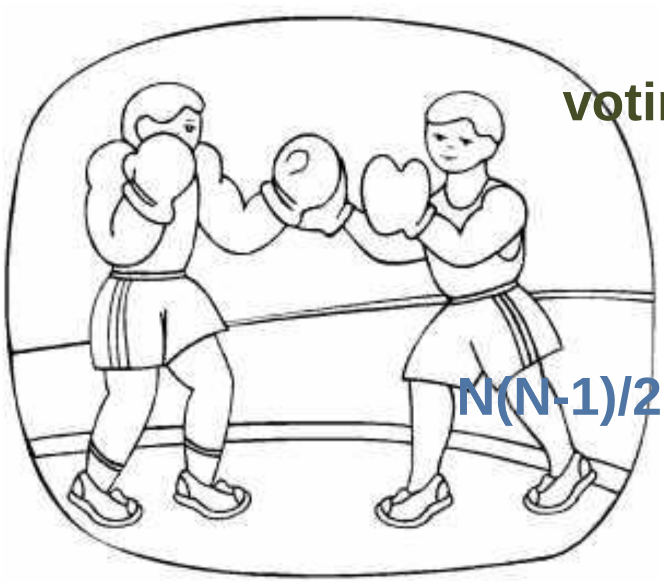
- Based on Binary Analysis
  - requires a multiclass analysis be broken down into series of binary classifications

  - One-Against-One
  - One-Against-All
  - DAGSVM

- "All-Together" Method
  - using decomposition
  - a one-shot multiclass classification needing a single optimization operation

Method based on Binary analysis

- **One-Against-One (1A1)**



- **also called MWV_SVM or Max-win voting SVM**

- involves constructing a learning system for each pair of classes

N(N-1)/2 machines

- When applied to a test data, each classification gives one vote to the winning class and the data is labeled with the class having most votes.

- If there are two identical votes, Max-win selects the class with the smallest index

Method based on Binary analysis

- **One-Against-All (1AA)**

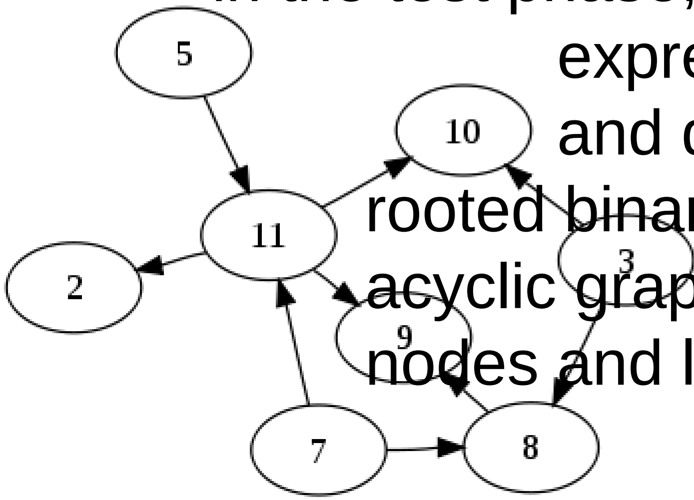
- - also called WTA_SVM or Winner-Takes-All method

- involves training N different binary classifiers, each one trained to distinguish the data in a single class from the data in all remaining classes.

When it is time to classify a new data, the N classifiers are run, and the classifier which outputs the largest (most positive) value is chosen.

METHOD BASED ON BINARY ANALYSIS

- Directed-Acylcic Graph SVM (DAG_SVM)
  - similar to 1A1
  - in the test phase, it uses a data structure to express decision node and decision algorithm (a rooted binary directed acyclic graph internal nodes and leaves)



"All-together" method

•Similar to the "One-Against-All" approach

•Mathematically one-step formula

•Is one-step more efficient?

NUMERICAL EXPERIMENTS

	One-against-one		DAG		One-against-all		[25], [27]		C&S	
Problem	C	rate	C	rate	C	rate	C	rate	C	rate
iris	2^4	**97.333**	2^8	**97.333**	2^{12}	96.000	2^5	**97.333**	2^0	87.333
wine	2^{-2}	**99.438**	2^{-2}	98.315	2^2	98.876	2^{-1}	98.876	2^{-1}	99.438
glass	2^8	**66.355**	2^4	63.551	2^5	58.879	2^9	65.421	2^6	62.617
vowel	2^5	**82.954**	2^6	81.439	2^{11}	50.000	2^8	67.424	2^6	63.068
vehicle	2^5	80.615	2^5	**80.851**	2^{12}	78.132	2^{10}	80.142	2^4	79.669
segment	2^{12}	**96.017**	2^{11}	95.844	2^{12}	93.160	2^8	95.454	2^{-2}	92.165

A COMPARISON USING THE LINEAR KERNEL (BEST RATES BOLD-FACED)

WINNER!

One-Against-One

DAG-SVM

Wait…how about Pairwise Coupling

- From a more recent study (2005)
- Combines posterior probabilities from individual One-Against-One binary classifications.
- Since output of SVM is a distance measurement, output must first be transformed into posterior probabilities.
- Pairwise coupling outperforms traditional binary SVM classifiers.

Hastie and Tibshirani

all the one-versus-one

binary classifiers

to obtain estimates of the

posterior probabilities $p_i = \mathrm{Prob}(\omega_i|x)$, $i = 1, . . . , M$

Kullbac-Liebler distance

$$l(p) = \sum_{i<j} n_{ij} \left(r_{ij} \log \frac{r_{ij}}{\mu_{ij}} + (1 - r_{ij}) \log \frac{1 - r_{ij}}{1 - \mu_{ij}} \right)$$

Platt's Sigmoid function

$$\mathrm{Prob}(\omega_1|\mathbf{x}) = \frac{1}{1+e^{Af+B}}$$

f: the output of the SVM associated with example x.

Parameters A and B: minimize the negative log-likelihood (NLL) function of the validation data.

Dataset	Training Set Size	Method			
		WTA_SVM	MWV_SVM	PWC_PSVM	PWC_KLR
ABE	280	1.92±0.65	1.96±0.65	1.16±0.63	1.85±0.59
	560	0.96±0.36	1.06±0.42	0.58±0.29	1.02±0.43
	1,120	0.46±0.20	0.50±0.24	0.34±0.17	0.57±0.26
DNA	300	10.15±1.26	9.87±0.90	9.23±1.73	9.73±0.75
	500	7.84±0.79	7.67±0.93	7.41±1.14	7.80±0.71
	1,000	5.59±0.39	5.72±0.57	5.50±0.69	5.76±0.54
SAT	1,000	11.07±0.58	11.03±0.73	10.27±0.92	11.20±0.55
	1,500	10.08±0.49	10.20±0.51	10.05±0.60	10.23±0.42
	2,000	9.51±0.31	9.61±0.39	9.47±0.65	9.66±0.37
SEG	250	9.43±0.54	7.97±1.23	6.66±2.24	7.54±1.24
	500	6.51±0.99	5.40±1.04	5.19±0.74	4.83±0.68
	1,000	4.89±0.71	4.35±0.79	4.08±0.52	3.96±0.68
WAV	150	17.21±1.37	17.75±1.39	13.20±3.70	15.59±1.13
	300	15.43±0.97	15.96±0.98	12.97±2.02	14.71±0.72
	600	14.09±0.55	14.56±0.80	13.47±1.09	13.81±0.41