Multiclass

Jordan Boyd-Graber
University of Colorado Boulder
LECTURE 14

Slides adapted from Rob Schapire and Fei Xia

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

.

Administrivia

- Feature Engineering HW Graded
- SVM HW
- Midterm in two weeks

Defining a Code Book

- Want to decide whether a name is German, Russian, or Chinese
- Using ECOC
- What do we need first?

- Want to decide whether a name is German, Russian, or Chinese
- Using ECOC
- What do we need first?

Class	b_1	b_2	b_3	b_4	b_5	b_6
Chinese	0	0	0	1	0	1
German	0	1	0	0	1	0
Russian	1	0	1	1	0	1
Argentine	1	1	1	0	1	0

Training Data

German	Russian	Chinese	Argentine
Mann	Dostoyevsky	Cao Xueqin	Puig
Goethe	Tolstoy	Lu Xun	Borges
Grass	Gogol	Gao Xingjian	Cortazar

German	Russian	Chinese	Argentine
Mann	Dostoyevsky	Cao Xueqin	Puig
Goethe	Tolstoy	Lu Xun	Borges
Grass	Gogol	Gao Xingjian	Cortazar

What are the training examples for each classifier?

German	Russian	Chinese	Argentine
Mann	Dostoyevsky	Cao Xueqin	Puig
Goethe	Tolstoy	Lu Xun	Borges
Grass	Gogol	Gao Xingjian	Cortazar

Class	b_1	b_2	b_3	b_4	b_5	b_6
Chinese	0	0	0	1	0	1
German	0	1	0	0	1	0
Russian	1	0	1	1	0	1
Argentine	1	1	1	0	1	0

What are the training examples for each classifier?

Training Examples

h_1 h_2 h_3 h_4 h_5 h_6

	h_1	h_2	h_3	h_4	h_5	h_6
Mann	-	+	-	-	+	-
Goethe	-	+	-	-	+	-
Grass	-	+	-	-	+	-

	h_1	h_2	h_3	h_4	h_5	h_6
Mann	-	+	-	-	+	-
Goethe	-	+	-	-	+	-
Grass	-	+	-	-	+	-
Dostoyevsky	+	-	+	+	-	+
Tolstoy	+	-	+	+	-	+
Gogol	+	-	+	+	-	+

Training Examples

	h_1	h_2	h_3	h_4	h_5	h_6
Mann	-	+	-	-	+	-
Goethe	-	+	-	-	+	-
Grass	-	+	-	-	+	-
Dostoyevsky	+	-	+	+	-	+
Tolstoy	+	-	+	+	-	+
Gogol	+	-	+	+	-	+
Cao Xue	-	-	-	+	-	+
Lu Xun	-	-	-	+	-	+
Gao Xingjian	-	-	-	+	-	+

	h_1	h_2	h_3	h_4	h_5	h_6
Mann	-	+	-	-	+	-
Goethe	-	+	-	-	+	-
Grass	-	+	-	-	+	-
Dostoyevsky	+	-	+	+	-	+
Tolstoy	+	-	+	+	-	+
Gogol	+	-	+	+	-	+
Cao Xue	-	-	-	+	-	+
Lu Xun	-	-	-	+	-	+
Gao Xingjian	-	-	-	+	-	+
Puig	+	+	+	-	+	-
Borges	+	+	+	-	+	-
Cortazar	+	+	+	-	+	-

What about h_2 vs. h_5?

Training Examples

	h_1	h_2	h_3	h_4	h_5	h_6
Mann	-	+	-	-	+	-
Goethe	-	+	-	-	+	-
Grass	-	+	-	-	+	-
Dostoyevsky	+	-	+	+	-	+
Tolstoy	+	-	+	+	-	+
Gogol	+	-	+	+	-	+
Cao Xue	-	-	-	+	-	+
Lu Xun	-	-	-	+	-	+
Gao Xingjian	-	-	-	+	-	+
Puig	+	+	+	-	+	-
Borges	+	+	+	-	+	-
Cortazar	+	+	+	-	+	-

What about h_2 vs. h_5? What about h_4 vs. h_2?

Classification

Class	b_1	b_2	b_3	b_4	b_5	b_6
Chinese	0	0	0	1	0	1
German	0	1	0	0	1	0
Russian	1	0	1	1	0	1
Argentine	1	1	1	0	1	0

-
-

Classification

Class	b_1	b_2	b_3	b_4	b_5	b_6
Chinese	0	0	0	1	0	1
German	0	1	0	0	1	0
Russian	1	0	1	1	0	1
Argentine	1	1	1	0	1	0

- $(0, 1, 0, 1, 1, 0) \rightarrow$
-

Classification

Class	b_1	b_2	b_3	b_4	b_5	b_6
Chinese	0	0	0	1	0	1
German	0	1	0	0	1	0
Russian	1	0	1	1	0	1
Argentine	1	1	1	0	1	0

- $(0, 1, 0, 1, 1, 0) \rightarrow$ German
-

Class	b_1	b_2	b_3	b_4	b_5	b_6
Chinese	0	0	0	1	0	1
German	0	1	0	0	1	0
Russian	1	0	1	1	0	1
Argentine	1	1	1	0	1	0

- $(0, 1, 0, 1, 1, 0) \rightarrow$ German
- $(0, 0, 0, 0, 0, 0) \rightarrow$

Class	b_1	b_2	b_3	b_4	b_5	b_6
Chinese	0	0	0	1	0	1
German	0	1	0	0	1	0
Russian	1	0	1	1	0	1
Argentine	1	1	1	0	1	0

- $(0, 1, 0, 1, 1, 0) \rightarrow$ German
- $(0, 0, 0, 0, 0, 0) \rightarrow$???

- Understand what your algorithm is doing when you ask it to multi class
- Features and training imbalance matter more than ever
- Debugging is often easier if **you** binarize the problem