Machine Learning Crash Course

Expectation Maximization

Dr. Henrich Matzinger and Dr. Greg Mayer
Machine Learning and AI Crash Course
December 2019

School of Mathematics
Georgia Institute of Technology, Atlanta, GA

Text Classification

Recall the document classification problem. We have **training data** with known topics.

$Topic$			dog	cat	car	home	run	be	nice
$T0$	Document1		1	0	0	1	1	0	0
$T0$	Document2		0	1	0	1	0	1	0
$T0$	Document3		1	0	0	0	0	1	0
$T1$	Document4		0	0	1	1	0	1	0
$T1$	Document5		0	0	1	0	1	0	1
$T1$	Document6		0	0	1	0	0	1	1

We calculate the conditional probability estimates:

$$T0: \quad \hat{\vec{p}} = (\frac{2}{3}, \frac{1}{3}, 0, \frac{2}{3}, \frac{1}{3}, \frac{2}{3}, 0), \qquad T1: \quad \hat{\vec{q}} = (0, 0, 1, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3})$$

We then use $\hat{\vec{p}}$ and $\hat{\vec{q}}$ to classify a new document.

Errors in Training Data Topics

- Assume that the topics are given to us but with 30% percent error.
- Can we still use the topics? How?

Example of partially messed up topic

Messed up Topic	True Topic			dog	cat	car	home	run	be	nice
$T0$	$T0$	Doc1		1	0	0	1	1	0	0
$T0$	$T0$	Doc2		0	1	0	1	0	1	0
$T0$	$T0$	Doc3		1	0	0	0	0	1	0
$T1$	$T1$	Doc4		0	0	1	1	0	1	0
$T1$	$T1$	Doc5		0	0	1	0	1	0	1
$T1$	$T1$	Doc6		0	0	1	0	0	1	1

Unknown estimated conditional probabilities:

$$T0: \quad \hat{\vec{p}} = (\frac{2}{3}, \frac{1}{3}, 0, \frac{2}{3}, \frac{1}{3}, \frac{2}{3}, 0) \qquad T1: \quad \hat{\vec{q}} = (0, 0, 1, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3})$$

What should we do?

First Step of Solution

Try an iterative approach.

Messed up Topic	True Topic			dog	cat	car	home	run	be	nice
$T0$	$T0$	Doc1		1	0	0	1	1	0	0
$T0$	$T0$	Doc2		0	1	0	1	0	1	0
$T0$	$T0$	Doc3		1	0	0	0	0	1	0
$T1$	$T1$	Doc4		0	0	1	1	0	1	0
$T1$	$T1$	Doc5		0	0	1	0	1	0	1
$T1$	$T1$	Doc6		0	0	1	0	0	1	1

Step 1: Use the messed up topic as if they were accurate.

$$T0: \quad \hat{\vec{p}}_I = (\frac{2}{3}, 0, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{1}{3}, \frac{1}{3}) \qquad T1: \quad \hat{\vec{q}}_I = (0, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, 0, 1, \frac{1}{3})$$

Step 2

Having estimated conditional probabilities with messed-up topics:

$$\hat{\vec{p}}_I = \left(\frac{2}{3}, 0, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right) \ , \ \hat{\vec{q}}_I = \left(0, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, 0, 1, \frac{1}{3}\right)$$

Classify every document in the data using $\hat{\vec{p}}_I$ and $\hat{\vec{q}}_I$ to create new estimated conditional probabilities:

$$\hat{\vec{p}}_{II}, \hat{\vec{q}}_{II}$$

Then iterate steps 1 and 2 until there is very little change in the topics.

Final Notes

- IF Naive bayse does not work with given true topics, this one will not work either $\Longrightarrow$ tricks are needed
- **Enormous potential**: if we have annotated only a small portion of the data ....