Naïve Bayes Classifier Application

Movie Reviews

Faisal Ahmad
Ben Cheng
Ivy Lee
Angela

April 22, 2013

Bayesian Inference

- *Motivation?*

Given that a patient tests positive, what is probability the patient is sick?

$$p(sick|+) = \frac{\overbrace{p(+|sick)}^{99/100}\ \overbrace{p(sick)}^{1/100}}{\underbrace{p(+)}_{99/100^2 + 99/100^2 \quad 198/100^2}} = \frac{99}{198} = \frac{1}{2}$$

where $p(+) = p(+|sick)\,p(sick) + p(+|healthy)\,p(healthy)$.

Model Representation

Representing each feature as Independent Bernoulli Random variable, probability of observing label x, given feature c is:

Why Naive?

- We assume features are independent

Maximum Likelihood

Under maximum likelihood inference we define the "best" parameter values as those for which the observed data are most probable:

$$\textit{Class (A)} = \arg max \log P(label\ i) + \sum_{j=1}^{n} \log[P(f_j \,|label\ i)]$$

Logarithmic Estimation and Smoothing

Logarithmic Estimation introduced due to:

- If training set is not fully representative, then for some $\mathbf{P(feature_j|Label_i)=0}$
- Possibility of floating points underflow.

Additive Smoothing:

- A constant smoothing factor is introduced-

POS Tagger

- Input: Documents and POS Type
- Tokenize the Documents into Words
- Tag the Words using nltk.pos_tag()
- Return only alpha character words with appropriate POS Type

Naïve Bayes Classifier

- Input: Training data, Boolean switch for log density and smoothing parameter
- Procedures
  - Clean Training Data
  - Train
  - Calculate (Smoothed) Posterior Probabilities
  - Classify (Most probable class)

Selecting features from training data

- Split reviews into 'positive' and 'negative'
  - Separately, search through the first 50 reviews, filtering out only adjectives/verbs using posTagger
  - From all words obtained, choose features from the top 10% most frequent words
- Combine chosen features from 'positive' and 'negative' training data
- Implement Naive Bayes classification with and without a smoothing parameter of 0.5
- Report error metrics (10 trials)
  - Hit (TP), False alarm (FP), Specificity (SPEC)

Results - Using adjectives

	Basic			Smooth (0.5)		
Experiment	**TP**	**FP**	**SPEC**	**TP**	**FP**	**SPEC**
1	77.8%	46.7%	53.3%	77.1%	45.9%	54.1%
2	74.8%	39.4%	60.6%	74.8%	39.0%	61.0%
…	…	…	…	…	…	…
9	66.2%	34.4%	65.6%	67.5%	33.5%	66.5%
10	62.9%	24.0%	76.0%	62.1%	23.4%	76.6%
Average	**71.5%**	**38.8%**	**61.2%**	**71.6%**	**38.3%**	**61.7%**

148 FEATURES

Results - Using adjectives

Top features	PosNegRatio	Bottom features	PosNeg Ratio
'accessible'	7.26	'bite'	0.19
'breathtaking'	6.57	'worst'	0.22
'ambitious'	5.18	'biblical'	0.26
'additional'	5.18	'bitchy'	0.26
'annual'	5.18	'awful'	0.30

Results - Including verbs

	Basic			Smooth (0.5)		
Experiment	**TP**	**FP**	**SPEC**	**TP**	**FP**	**SPEC**
1	68.2%	39.0%	61.0%	68.6%	38.6%	61.4%
2	81.1%	54.0%	46.0%	81.1%	53.8%	46.2%
…	…	…	…	…	…	…
9	73.2%	46.1%	53.9%	73.4%	46.1%	53.9%
10	68.8%	44.1%	55.9%	68.8%	43.5%	56.5%
Average	**68.7%**	**44.3%**	**55.7%**	**68.8%**	**44.1%**	**55.9%**

116 FEATURES

Results - Including verbs

Top features	PosNeg Ratio	Bottom features	PosNeg Ratio
'disturbing'	3.05	'dressed'	0.30
'opened'	2.96	'dumb'	0.36
'minor'	2.50	'pull'	0.51
'compelling'	2.20	'guess'	0.53
'loose'	1.88	'replaced'	0.56

Conclusion

- Smoothing effect not significant
- Using only adjectives attains better result than combining adjectives and verbs.