Martin Mundt, Dr. Iuliia Pliushch, Prof. Dr. Visvanathan Ramesh

Pattern Analysis & Machine Intelligence Praktikum: MLPR WS-19/20

Week 4: Naive Bayes

Bayes’ rule

Likelihood

Class Prior Probability

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Predictor Prior Probability

Posterior Probability

https://www.thelearningmachine.ai/naive

Bayes’ rule: the naive assumption

- one feature case: $P(C|X) \propto P(X|C)P(C)$

- several features case: $P(C|X_1, X_2 \ldots X_n) \propto P(X_1, X_2 \ldots X_n|C)P(C)$

- the independency assumption: $P(C|X_1, X_2 \ldots X_n) \propto P(X_1|C)P(X_2|C) \ldots P(X_n|C)P(C)$

Naive Bayes example

Feature 1	*Feature 2*	*Classification*
Weather	**Wind**	**Play**
Sunny	Strong	No
Overcast	Weak	Yes
Rainy	Moderate	Yes
Sunny	Weak	Yes
Sunny	Weak	Yes
Overcast	Weak	Yes
Rainy	Strong	No
Rainy	Strong	No
Sunny	Weak	Yes
Rainy	Weak	Yes
Sunny	Strong	No
Overcast	Weak	Yes
Overcast	Weak	Yes
Rainy	Moderate	No

$$P(C=yes|X_1=Moderate, X_2=Sunny) \propto$$
$$P(X_1=Moderate, X_2=Sunny|C=yes)P(C=yes)=$$
$$P(X_1=Moderate|C=yes)P(X_2=Sunny|C=yes)P(C=yes)=$$
$$\frac{1}{14}\frac{3}{14}\frac{9}{14}=0.0098$$

$$P(C=no|X_1=Moderate, X_2=Sunny) \propto$$
$$P(X_1=Moderate, X_2=Sunny|C=no)P(C=no)=$$
$$P(X_1=Moderate|C=no)P(X_2=Sunny|C=no)P(C=no)=$$
$$\frac{1}{14}\frac{2}{14}\frac{5}{14}=0.0036$$

$$P(C=yes|X_1=Moderate, X_2=Sunny) > P(C=no|X_1=Moderate, X_2=Sunny)$$

https://www.thelearningmachine.ai/naive

Bag-of-word representation

- **n**: number of words in a document
- **m**: vocabulary size
- $\mathbf{x}_j$: frequency of a word j in a document
- θ_j : probability of a word j

$$P(x;\theta)=\frac{n!}{\prod_{j=1..m} x_j!}\prod_{j=1..m}\theta_j^{x_j}$$

http://cseweb.ucsd.edu/~elkan/250B/topicmodels.pdf

Naive Bayes: text case

- **Training data**: a corpus of messages/texts
- **Task**: classify messages as spam or non-spam
- **Attributes**: words in the messages
- **Naive Bayes classifier**:

$$c_{NB} = \arg\max_{c_i \in spam, non-spam} P(c_i) \prod P(w_i | c_i)$$

http://www.coli.uni-saarland.de/~crocker/Teaching/Connectionist/lecture10_4up.pdf

Multinomial Naive Bayes: text case

- **vocabulary V**: set of all words in the training corpus

- **N**: number of documents in the training corpus
- $\mathbf{N_{ci}}$: number of documents with class c_i

- $\mathbf{n_k}$: number of times (frequency) word w_k occurs in texts of class $\mathbf{c_i}$
- $\mathbf{n_{ci}}$: number of words in texts of class $\mathbf{c_i}$

$$P(w_k|c_i)=\frac{n_k+1}{n_{ci}+|V|} \qquad P(c_i)=\frac{N_{ci}}{N}$$

https://www.inf.ed.ac.uk/teaching/courses/inf2b/learnnotes/inf2b-learn07-notes-nup.pdf

Naive Bayes: text case example

	Doc	Words	Class
Training	1	Chinese Beijing Chinese	c
	2	Chinese Chinese Shanghai	c
	3	Chinese Macao	c
	4	Tokyo Japan Chinese	j
Test	5	Chinese Chinese Chinese Tokyo Japan	?

$$P(w_k|c_i) = \frac{count(w_k, c_i) + 1}{count(c_i) + |V|}$$

$$P(c_i) = \frac{N_{ci}}{N}$$

$$P(C) = 3/4$$

$$P(Chinese|C) = \frac{(5+1)}{(8+6)} = 6/14 = 3/7$$

$$P(Chinese|J) = \frac{(1+1)}{(3+6)} = 2/9$$

$$P(Tokyo|C) = \frac{(0+1)}{(8+6)} = 1/14$$

https://web.stanford.edu/class/cs124/lec/naivebayes.pdf