# Artificial Intelligence

Unit 06
Naïve Bayes

By:
Syeda Saleha Raza

AL NAFI,
A company with a focus on education,
wellbeing and renewable energy.

اَللّٰهُمَّ إِنِّي أَسْأَلُكَ عِلْمًا نَّافِعًا،

وَرِزْقًا طَيِّبًا، وَعَمَلًا مُّتَقَبَّلًا،

(O Allah, I ask You for beneficial knowledge, goodly provision and acceptable deeds)

اے اللہ ، میں آپ سے سوال کرتی ہوں نفع بخش علم کا، طیّب رزق کا، اور اس عمل کا

(Sunan Ibn Majah: 925)

# Outline

- Naïve Bayes Model
  - Probabilities - Recap
  - Learning Naïve Bayes
  - Strength and Weaknesses
  - Code Demo

# Naïve Bayes

# Bayes' Rule Applied to Documents and Classes

A B

$$\boxed{P(A/B) = \frac{P(B/A)P(A)}{P(B)}}$$

For a document *d* and a class *c*

$$P(c \mid d) = \frac{P(d \mid c)P(c)}{P(d)}$$

# Parameter Estimation

P( 'amazing' | 'positive')

= 1/2 = 0.5

$$\hat{P}(w_i \mid c_j) = \frac{count(w_i, c_j)}{\sum_{w \in V} count(w, c_j)}$$

fraction of times word $w_i$ appears among all words in documents of topic $c_j$

| Review | Class |
|---|---|
| "The movie was amazing!" | Positive |
| "I didn't like the film." | Negative |
| "Great acting and an engaging plot." | Positive |
| "Terrible movie, I would not recommend it to anyone." | Negative |

Vocab

# Applying Multinomial Naive Bayes Classifiers to Text Classification

$$c_{NB} = \underset{c_j \in C}{\operatorname{argmax}}\, P(c_j) \prod_{i \in positions} P(x_i \mid c_j)$$

 

# Naive Bayes - Example

| Review | Class |
|---|---|
| "The movie was amazing!" | Positive |
| "I didn't like the film." | Negative |
| "Great acting and an engaging plot." | Positive |
| "Terrible movie, I would not recommend it to anyone." | Negative |

# Naive Bayes - Example

| Review | Class |
| --- | --- |
| "The movie was amazing!" | Positive |
| "I didn't like the film." | Negative |
| "Great acting and an engaging plot." | Positive |
| "Terrible movie, I would not recommend it to anyone." | Negative |

[“the,” “movie,” “was,” “amazing,” “I,” “didn’t,” “like,” “film,” “great,” “acting,” “and,” “an,” “engaging,” “plot,” “terrible,” “would,” “not,” “recommend,” “it,” “to,” “anyone”].

# Naive Bayes - Example

| Review | Class |
|---|---|
| "The movie was amazing!" | Positive |
| "I didn't like the film." | Negative |
| "Great acting and an engaging plot." | Positive |
| "Terrible movie, I would not recommend it to anyone." | Negative |

[“the,” “movie,” “was,” “amazing,” “I,” “didn’t,” “like,” “film,” “great,” “acting,” “and,” “an,” “engaging,” “plot,” “terrible,” “would,” “not,” “recommend,” “it,” “to,” “anyone”].

| Review | BoW Representation | Class |
|---|---|---|
| "The movie was amazing!" | [1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0] | Positive |
| "I didn't like the film." | [0, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0] | Negative |
| "Great acting and an engaging plot." | [0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0] | Positive |
| "Terrible movie, I would not recommend it to anyone." | [0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1] | Negative |

# Naive Bayes - Example

- Now, let's assume we have a new movie review:
  - "The acting was great!"

+ve

- We can convert this review into its BoW representation:
  - [0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0].

$$\boxed{c_{NB} = \underset{c \in C}{\operatorname{argmax}} P(c_j) \prod_{x \in X} P(x \mid c)}$$

| Class | Prior Probability | Likelihood | Posterior Probability |
|---|---|---|---|
| Positive | 0.5 | 0.004 | 0.002 |
| Negative | 0.5 | 0.000000000002 | 0.000000000001 |

# Classification using Naïve Bayes

$$\hat{P}(c) = \frac{N_c}{N}$$

$$\hat{P}(w \mid c) = \frac{count(w,c)+1}{count(c)+|V|}$$

| | 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 | ? |

c

j

**Priors:**

$P(c) = \frac{3}{4}$

$P(j) = \frac{1}{4}$

**Conditional Probabilities:**

P(Chinese|*c*) = (5+1) / (8+6) = 6/14 = 3/7

P(Tokyo|*c*) = (0+1) / (8+6) = 1/14

P(Japan|*c*) = (0+1) / (8+6) = 1/14

P(Chinese|*j*) = (1+1) / (3+6) = 2/9

P(Tokyo|*j*) = (1+1) / (3+6) = 2/9

P(Japan|*j*) = (1+1) / (3+6) = 2/9

$P(w_i / c_i)$

**Choosing a class:**

P(c|d5) ∝ 3/4 * $(3/7)^3$ * 1/14 * 1/14 ≈ 0.0003

P(j|d5) ∝ 1/4 * $(2/9)^3$ * 2/9 * 2/9 ≈ 0.0001

41

[7_NB.pdf (stanford.edu)]

# Dealing with small values

- Multiplying lots of small probabilities (all are under 1) can
- lead to numerical underflow …

$$\log \prod_i x_i = \sum_i \log x_i$$

# Smoothing

- What if we have seen no training documents with the word fantastic and classified in the topic positive (thumbs-up)
- To deal with low counts, it can be helpful to smooth probabilities
- Smoothing term ! is a hyperparameter, which must be tuned on a development set
- **Laplace (add-1) smoothing: widely used**

# Strength

- Very Fast, low storage requirements
- Robust to Irrelevant Features Irrelevant Features cancel each other without affecting results
- Very good in domains with many equally important features Decision Trees suffer from fragmentation in such cases – especially if little data
- Optimal if the independence assumptions hold: If assumed independence is correct, then it is the Bayes Optimal Classifier for problem
- A good dependable baseline for text classification (But we will see other classifiers that give better accuracy)

# Naïve Bayes - Code Walkthrough

# Resources

- https://www.coursera.org/learn/machine-learning/lecture/du981/backpropagation-intuition
- https://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example/
- https://scikit-neuralnetwork.readthedocs.io/en/latest/index.html
- 7_NB.pdf (stanford.edu)

# جزاك الله

To ask questions, Please use communities link (for respective course) within portal

https://portal.alnafi.com/enrollments