Классификация.

предсказываем значение из конечного мн-ва

1) {спам/не спам}
2) {хорошо/плохо} тональность
3) {спорт, политика, погода, ...}

Классифицируем объекты (тексты)

объект → признак1 = ... признак2 = ... признак = ...

↑ бинар. значения (для наивн. байеса)

⇒ класс = ...

В текстах для метода NB обычно так: для каждого слова из словаря указывается, есть оно или нет:

Текст "мне понравилось" text1

"мне не очень" text2

V=словарь = {мне, не, очень, понравилось}

text1: {мне = True; не = False, очень = False, понравилось = True}

text2: {мне = True не = True очень = True понравилось = False}

Такое представление текстов называется

Bag of Words — мешок слов.

(нет информации о порядке слов)

можно хранить информацию в виде таблицы:

	мне не он. нравится
text1	1 0 0 1
text2	1 1 1 0

$\begin{pmatrix}1\\2\end{pmatrix}$ — 1: класс 1, positive; 2: класс 2, negative

строки - объекты (тексты)
столбцы - признаки

если есть информация для обучения, правильные ответы, то называется еще меткой

после обучения мы даём модели text3, в виде набора признаков и получаем предсказание класса (positive / negative)

Метод: будем определять вероятность класса по набору признаков

$$P\{C = positive \mid W_1 = w_1 \ \ W_2 = w_2 \ \dots \ W_N = w_n\}$$

C — класс; W_1 — feature, признак, слово; w_1 — один 1; w_2 — ...

Сокращенно:

$$P\{pos \mid w_1, w_2 \dots w_n\} =$$

$$= P\{w_1, w_2 \dots w_n \mid pos\} \cdot \frac{P(pos)}{P(w_1 \dots w_n)}$$

Формула Байеса

$$P(A \mid B) \overset{опр}{=} \frac{P(AB)}{P(B)}$$

$$P(B \mid A) \overset{опр}{=} \frac{P(BA)}{P(A)}$$

$$\Downarrow$$

Формула: $P(A \mid B) = P(B \mid A) \cdot \frac{P(A)}{P(B)}$

$$P\{neg \mid w_1 \dots w_n\} =$$

$$P\{w_1 \dots w_n \mid neg\} \cdot \frac{P(neg)}{P(w_1 \dots w_n)}$$

$P\{pos \dots\} \overset{?}{>} P\{neg \dots\}$

$P(w_1 \dots w_n)$ — сокращаются.

Итого, чтоб выбрать класс, достаточно сравнить

$P\{w_1 \dots w_n | pos\} \cdot P\{pos\}$ и

$P\{w_1 \dots w_n | neg\} \cdot P\{neg\}$

$P(AB) = P(A)P(B)$ если A, B независимы

– предположение о наивности:

$P\{w_1 \dots w_n | c\} = P\{w_1 | c\} \dots P\{w_n | c\}$

итого, чтоб определить класс, надо сравнить

$P\{w_1 | pos\} \dots P\{w_n | pos\} \cdot P\{pos\}$ и

$P\{w_2 | neg\} \dots P\{w_n | neg\} \cdot P\{neg\}$

каждый множитель можно оценить по частоте в корпусе

$$P\{W_i = w_i | C = c\} = \frac{\text{c(документы класса c, где } W_i \text{ есть или нет)}}{\text{c(документы класса c)}}$$

pos, neg

Пример $P\{C = c\} = \frac{\text{док из } C}{\text{всего док}}$

	тлс	не	оч.	понр.	
text1	1	0	0	1	pos
text2	1	1	1	0	neg
text3	0	0	1	1	pos

$P\{тлс = 1 | C = pos\} = \frac{1 - \text{1 pos док со словом тлс}}{2 - \text{pos док.}}$

$P\{тлс = 0 | C = pos\} = \frac{1}{2}$

$P\{\text{"понравилось"} | C = pos\} = \frac{2}{2}$ — 2 документа из pos со словом "понравилось"

$P(\text{не}=1 \mid c=\text{пед}) = \frac{1}{1}$ — док. со словом не / — всего пед док.

Вероятности классов:

$P(C=\text{пед}) = 1/3$ $\quad P(C=\text{рос}) = 2/3$

Как и в прошлых моделях $P(\dots) = 0$ бывают $=0$. Тоже нужно сглаживание — чтоб все вероятности $\neq 0$.

На практике, вычисляя

$P(w_1 \mid c) \dots P(w_n \mid c) P(c) \nearrow \max$

берут $-\log$:

$-\log P(w_1 \mid c) - \dots - \log P(w_n \mid c) - \log P(c)$

это должно быть меньше.

штраф за слова — штраф за класс

Что еще можно делать после обучения:

можно найти самые большие значения

$P(w_i \mid c)$ — w_i сильно говорит в пользу класса c