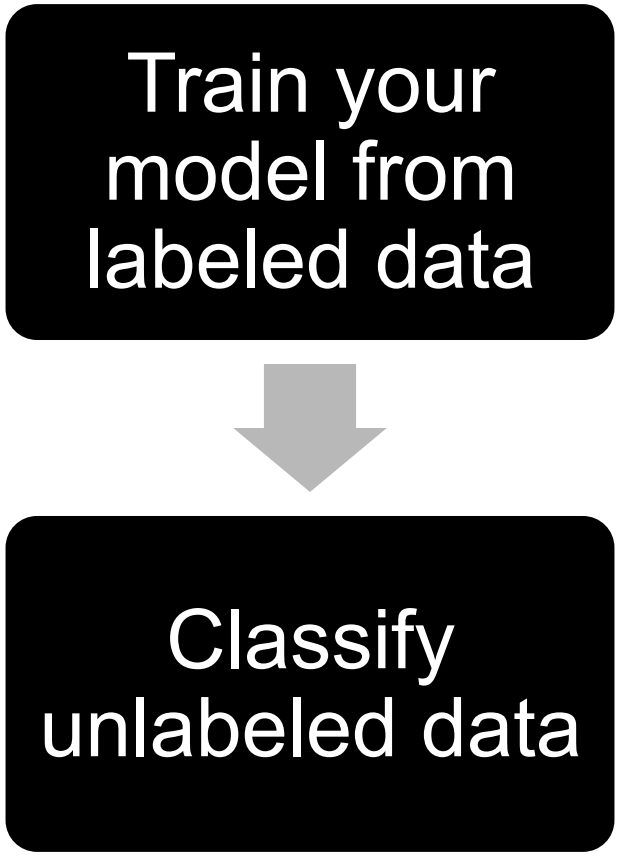
Lab3 Part3 Assignment

Given an email, predict whether it is spam or not



Key idea

- Make a probabilistic model of data within each category

- Why we call it “Naive Bayes”
  - Assume each feature is independent of other features
  - Too simple, sometimes naive?
  - Many times perform surprisingly well

Spam/Ham Classification

- **Features: bag-of-words (all words with counts but without accounting for the order)**
- **For each class ck compute P(ck|bag-of-words) and pick the class with the highest probability**

Probability Basics

Prior probability: $P(X)$

Conditional probability: $P(X_1 \mid X_2), P(X_2 \mid X_1)$

Joint probability: $\mathbf{X} = (X_1, X_2), P(\mathbf{X}) = P(X_1, X_2)$

Relationship: $P(X_1, X_2) = P(X_2 \mid X_1)P(X_1) = P(X_1 \mid X_2)P(X_2)$

Independence: $P(X_2 \mid X_1) = P(X_2), P(X_1 \mid X_2) = P(X_1), P(X_1, X_2) = P(X_1)P(X_2)$

- **Bayesian Rule:**

$$P(C \mid \mathbf{X}) = \frac{P(\mathbf{X} \mid C)P(C)}{P(\mathbf{X})} \qquad Posterior = \frac{Likelihood \times Prior}{Evidence}$$

Naive Bayes Text Classification

- **Given a document d, what class does it belong to?**
- **Find the most likely class *c↓pred***

$$c_{pred} = \arg\max_{c_k} P(c_k \mid d)$$

$$= \arg\max_{c_k} \frac{P(c_k)P(d \mid c_k)}{P(d)}$$

$$= \arg\max_{c_k} \frac{P(c_k)P(d \mid c_k)}{\sum_{k=1}^{K} P(c_k)P(d \mid c_k)}$$

Naive Bayes Text Classification

$$c_{pred} = \arg\max_{c_k} \frac{P(c_k)P(d \mid c_k)}{\sum_{k=1}^{K} P(c_k)P(d \mid c_k)}$$

- **If document d is L words long**

$$P(d \mid c_k) = P(w_1 \mid c_k)\, P(w_2 \mid c_k)\, P(w_3 \mid c_k) \ldots P(w_L \mid c_k)$$

- **Note: the denominator (in the equation on the top) is the same for all classes and omitting it will not affect the comparison of classes**

An Example: Training

Documents with labels

SPAM	click for pharmacy
¬SPAM	free time today
SPAM	online pharmacy link
¬SPAM	no free time
¬SPAM	free good pharmacy
SPAM	pharmacy free link
¬SPAM	for time today
¬SPAM	time is money

Vocabulary (12 words in total)

click
for
pharmacy
free
time
today
online
link
no
good
is
money

An Example: Training

Documents with labels

SPAM	click for pharmacy
¬SPAM	free time today
SPAM	online pharmacy link
¬SPAM	no free time
¬SPAM	free good pharmacy
SPAM	pharmacy free link
¬SPAM	for time today
¬SPAM	time is money

Vocabulary size: 12

P(spam) = 3/8

P(¬spam) = 5/8

$$P(c_k) = \frac{count(c_k)}{M}$$

count(c_k): number of documents of class c_k

M: total number of documents

An Example: Training

Documents with labels

SPAM	click for pharmacy
¬SPAM	free time today
SPAM	online pharmacy link
¬SPAM	no free time
¬SPAM	free good pharmacy
SPAM	pharmacy free link
¬SPAM	for time today
¬SPAM	time is money

Vocabulary size: 12

P(spam) = 3/8

P(¬spam) = 5/8

P(pharmacy|spam) = 1/3

$$P(w_l \mid c_k) = \frac{count(w_l, c_k)}{count(w, c_k)}$$

count(w_l, c_k): number of times the word w_l appears in documents of class c_k

count(w, c_k): total number of words in documents of class c_k

An Example: Classification

Msg = "pharmacy for pharmacy"
Classify Msg as spam or ¬spam

$$c_{pred} = \arg\max_{c_k} P(c_k \mid d) = \arg\max_{c_k} \frac{P(c_k)P(d \mid c_k)}{\sum_{k=1}^{K} P(c_k)P(d \mid c_k)}$$

$$P(spam \mid Msg) = \frac{P(spam)P(Msg \mid spam)}{P(spam)P(Msg \mid spam) + P(\neg spam)P(Msg \mid \neg spam)}$$

$$P(\neg spam \mid Msg) = \frac{P(\neg spam)P(Msg \mid \neg spam)}{P(spam)P(Msg \mid spam) + P(\neg spam)P(Msg \mid \neg spam)}$$

if P(spam|Msg) > P(¬spam|Msg) then Msg is classified as spam
else if P(spam|Msg) < P(¬spam|Msg) then Msg is classified as ¬ spam
else cannot decide
Note: the denominator is the same for all classes and omitting it will not affect the comparison of classes

An Example: Classification

Documents with labels

SPAM	click for pharmacy
¬SPAM	free time today
SPAM	online pharmacy link
¬SPAM	no free time
¬SPAM	free good pharmacy
SPAM	pharmacy free link
¬SPAM	for time today
¬SPAM	time is money

Vocabulary size: 12

P(spam) = 3/8

P(¬spam) = 5/8

P(pharmacy|spam) = 1/3

P(pharmacy|¬spam) = 1/15

Msg = "pharmacy for pharmacy"

$$P(spam \mid Msg) = \frac{P(spam)P(Msg \mid spam)}{P(spam)P(Msg \mid spam) + P(\neg spam)P(Msg \mid \neg spam)}$$

An Example: Classification

Documents with labels

SPAM	click for pharmacy
¬SPAM	free time today
SPAM	online pharmacy link
¬SPAM	no free time
¬SPAM	free good pharmacy
SPAM	pharmacy free link
¬SPAM	for time today
¬SPAM	time is money

Vocabulary size: 12

P(spam) = 3/8

P(¬spam) = 5/8

P(pharmacy|spam) = 1/3

P(pharmacy|¬spam) = 1/15

Msg = "pharmacy for pharmacy"

$$P(spam \mid Msg) = \frac{\frac{3}{8} P(Msg \mid spam)}{\frac{3}{8} P(Msg \mid spam) + \frac{5}{8} P(Msg \mid \neg spam)}$$

An Example: Classification

Documents with labels

SPAM	click for pharmacy
¬SPAM	free time today
SPAM	online pharmacy link
¬SPAM	no free time
¬SPAM	free good pharmacy
SPAM	pharmacy free link
¬SPAM	for time today
¬SPAM	time is money

Vocabulary size: 12

P(spam) = 3/8

P(¬spam) = 5/8

P(pharmacy|spam) = 1/3

P(pharmacy|¬spam) = 1/15

P(for|spam)=1/9

P(for|¬spam)=1/15

Msg = "pharmacy for pharmacy"

$$P(spam \mid Msg) = \frac{\frac{3}{8} P(Msg \mid spam)}{\frac{3}{8} P(Msg \mid spam) + \frac{5}{8} P(Msg \mid \neg spam)}$$

$$P(Msg \mid spam) = P(pharmacy \mid spam) P(for \mid spam) P(pharmacy \mid spam)$$

$$= \frac{1}{3} \times \frac{1}{9} \times \frac{1}{3} = \frac{1}{81}$$

An Example: Classification

Documents with labels

SPAM	click for pharmacy
¬SPAM	free time today
SPAM	online pharmacy link
¬SPAM	no free time
¬SPAM	free good pharmacy
SPAM	pharmacy free link
¬SPAM	for time today
¬SPAM	time is money

Vocabulary size: 12

P(spam) = 3/8

P(¬spam) = 5/8

P(pharmacy|spam) = 1/3

P(pharmacy|¬spam) = 1/15

P(for|spam)=1/9

P(for|¬spam)=1/15

Msg = "pharmacy for pharmacy"

$$P(spam \mid Msg) = \frac{\frac{3}{8} \times \frac{1}{81}}{\frac{3}{8} \times \frac{1}{81} + \frac{5}{8} P(Msg \mid \neg spam)}$$

An Example: Classification

Documents with labels

SPAM	click for pharmacy
¬SPAM	free time today
SPAM	online pharmacy link
¬SPAM	no free time
¬SPAM	free good pharmacy
SPAM	pharmacy free link
¬SPAM	for time today
¬SPAM	time is money

Vocabulary size: 12

P(spam) = 3/8

P(¬spam) = 5/8

P(pharmacy|spam) = 1/3

P(pharmacy|¬spam) = 1/15

P(for|spam)=1/9

P(for|¬spam)=1/15

Msg = "pharmacy for pharmacy"

$$P(spam \mid Msg) = \frac{\frac{1}{216}}{\frac{1}{216} + \frac{5}{8} P(Msg \mid \neg spam)}$$

An Example: Classification

Documents with labels

SPAM	click for pharmacy
¬SPAM	free time today
SPAM	online pharmacy link
¬SPAM	no free time
¬SPAM	free good pharmacy
SPAM	pharmacy free link
¬SPAM	for time today
¬SPAM	time is money

Vocabulary size: 12

P(spam) = 3/8

P(¬spam) = 5/8

P(pharmacy|spam) = 1/3

P(pharmacy| ¬ spam) = 1/15

P(for|spam)=1/9

P(for| ¬ spam)=1/15

Msg = "pharmacy for pharmacy"

$$P(spam \mid Msg) = \frac{\frac{1}{216}}{\frac{1}{216} + \frac{5}{8} P(Msg \mid \neg spam)}$$

$$P(Msg \mid \neg spam) = P(pharmacy \mid \neg spam) P(for \mid \neg spam) P(pharmacy \mid \neg spam)$$

$$= \frac{1}{15} \times \frac{1}{15} \times \frac{1}{15} = \frac{1}{3375}$$

An Example: Classification

Documents with labels

SPAM	click for pharmacy
¬SPAM	free time today
SPAM	online pharmacy link
¬SPAM	no free time
¬SPAM	free good pharmacy
SPAM	pharmacy free link
¬SPAM	for time today
¬SPAM	time is money

Vocabulary size: 12

P(spam) = 3/8

P(¬spam) = 5/8

P(pharmacy|spam) = 1/3

P(pharmacy|¬spam) = 1/15

P(for|spam)=1/9

P(for|¬spam)=1/15

Msg = “pharmacy for pharmacy”

$$P(spam \mid Msg) = \frac{\frac{1}{216}}{\frac{1}{216} + \frac{5}{8} \times \frac{1}{3375}}$$

An Example: Classification

Documents with labels

SPAM	click for pharmacy
¬SPAM	free time today
SPAM	online pharmacy link
¬SPAM	no free time
¬SPAM	free good pharmacy
SPAM	pharmacy free link
¬SPAM	for time today
¬SPAM	time is money

Vocabulary size: 12

P(spam) = 3/8

P(¬spam) = 5/8

P(pharmacy|spam) = 1/3

P(pharmacy|¬spam) = 1/15

P(for|spam)=1/9

P(for|¬spam)=1/15

Msg = “pharmacy for pharmacy”

P(spam|Msg) = 25/26

What happens if Msg = “time for pharmacy”?

Add-one Smoothing

Documents with labels

SPAM	click for pharmacy
¬SPAM	free time today
SPAM	online pharmacy link
¬SPAM	no free time
¬SPAM	free good pharmacy
SPAM	pharmacy free link
¬SPAM	for time today
¬SPAM	time is money

Vocabulary size: 12

P(spam) = 3/8

P(¬spam) = 5/8

P(pharmacy|spam) = 1/3

P(pharmacy|¬spam) = 1/15

P(for|spam)=1/9

P(for|¬spam)=1/15

P(time|spam)=0

P(time|¬spam)=4/15

Msg = "time for pharmacy"

$$P(spam \mid Msg) = \frac{\frac{3}{8}P(Msg \mid spam)}{\frac{3}{8}P(Msg \mid spam) + \frac{5}{8}P(Msg \mid \neg spam)}$$

$$P(Msg \mid spam) = P(time \mid spam)P(for \mid spam)P(pharmacy \mid spam)$$
$$= 0 \times \frac{1}{9} \times \frac{1}{3} = 0$$

Add-one Smoothing

Documents with labels

SPAM	click for pharmacy
¬SPAM	free time today
SPAM	online pharmacy link
¬SPAM	no free time
¬SPAM	free good pharmacy
SPAM	pharmacy free link
¬SPAM	for time today
¬SPAM	time is money

Vocabulary size: 12

P(spam) = 3/8

P(¬spam) = 5/8

P(pharmacy|spam) = 1/3

P(pharmacy|¬spam) = 1/15

P(for|spam)=1/9

P(for|¬spam)=1/15

P(time|spam)=0

P(time|¬spam)=4/15

Msg = "time for pharmacy"

$$P(spam \mid Msg) = \frac{\frac{3}{8} \times 0}{\frac{3}{8} \times 0 + \frac{5}{8} P(Msg \mid \neg spam)}$$

Add-one Smoothing

Computing $P(c_k)$, e.g., P(spam) or P(¬spam)

Same formula with or without smoothing (we assume that we have enough documents in our training data for each class so no smoothing is required)

$$P(c_k) = \frac{count(c_k)}{M}$$

count(c_k): number of documents of class c_k
M: total number of documents

Documents with labels

SPAM	click for pharmacy
¬SPAM	free time today
SPAM	online pharmacy link
¬SPAM	no free time
¬SPAM	free good pharmacy
SPAM	pharmacy free link
¬SPAM	for time today
¬SPAM	time is money

Vocabulary size: 12
P(spam) = 3/8
P(¬spam) = 5/8

Add-one Smoothing

Computing $P(w_l|c_k)$, e.g., P(pharmacy|spam) or P([pharmacy|¬spam)

Without smoothing

$$P(w_l \mid c_k) = \frac{count(w_l, c_k)}{count(w, c_k)}$$

With smoothing

$$P(w_l \mid c_k) = \frac{count(w_l, c_k)+1}{count(w, c_k)+V}$$

count(w_l, c_k): number of times the word w_l appears in documents of class c_k
count(w, c_k): total number of words in documents of class c_k
V: vocabulary size

Documents with labels

SPAM	click for pharmacy
¬SPAM	free time today
SPAM	online pharmacy link
¬SPAM	no free time
¬SPAM	free good pharmacy
SPAM	pharmacy free link
¬SPAM	for time today
¬SPAM	time is money

Vocabulary size: 12
P(spam) = 3/8
P(¬spam) = 5/8

Without smoothing:

P(pharmacy|spam) = 1/3
P(time|spam) = 0

With smoothing:

P(pharmacy|spam)=(3+1)/(9+12)=4/21
P(time|spam)=(0+1)/(9+12)=1/21

Add-one Smoothing

Documents with labels

SPAM	click for pharmacy
¬SPAM	free time today
SPAM	online pharmacy link
¬SPAM	no free time
¬SPAM	free good pharmacy
SPAM	pharmacy free link
¬SPAM	for time today
¬SPAM	time is money

Vocabulary size: 12

P(spam) = 3/8

P(¬spam) = 5/8

With smoothing:

P(pharmacy|spam) = 4/21

P(pharmacy|¬spam) = 2/27

P(for|spam)=2/21

P(for|¬spam)=2/27

P(time|spam)=1/21

P(time|¬spam)=5/27

Msg = "time for pharmacy"

$$P(spam \mid Msg) = \frac{\frac{3}{8} P(Msg \mid spam)}{\frac{3}{8} P(Msg \mid spam) + \frac{5}{8} P(Msg \mid \neg spam)}$$

$$P(Msg \mid spam) = P(time \mid spam) P(for \mid spam) P(pharmacy \mid spam)$$

$$= \frac{1}{21} \times \frac{2}{21} \times \frac{4}{21} = \frac{8}{9261}$$