Lecture 9: Naïve Bayes Classifier

COMS10014 Mathematics for Computer Science A

`cs-uob.github.io/COMS10014/ and github.com/coms10011/2020_21`

November 2020

Machine learning and probabilities

Many learning algorithms can be thought of as machines for estimating probabilities, often in the face of insufficient data to estimate the probabilities required.

Spam filter

$$W = (\text{enlargement}, \text{xxx}, \text{cheapest}, \text{pharmaceuticals}, \text{satisfied}, \text{leeds})$$

Spam filter

$$W = (\text{enlargement}, \text{xxx}, \text{cheapest}, \text{pharmaceuticals}, \text{satisfied}, \text{leeds})$$

Spam filter

$$W = (\text{enlargement}, \text{xxx}, \text{cheapest}, \text{pharmaceuticals}, \text{satisfied}, \text{leeds})$$

Binary vector

Say $\mathbf{w}$ is a vector of zeros and ones indicating the presence or absence of different potential spam words in an email.

$$\mathbf{w} = (1, 1, 0, 0, 0, 1)$$

for

$$W = (\text{enlargement}, \text{xxx}, \text{cheapest}, \text{pharmaceuticals}, \text{satisfied}, \text{leeds})$$

Some notation

Now let S represent the event of an email being spam.

$$P(S|\mathbf{w}) > T$$

Counting

$$P(S|(1,1,0,0,0,1)) = \frac{\#\{\text{spam with enlargement, xxx and leeds}\}}{\#\{\text{all emails with enlargement, xxx and leeds}\}}$$

However there are $2^6 = 64$ different ws!

Bayes

$$P(S|\mathbf{w}) = \frac{P(\mathbf{w}|S)P(S)}{P(\mathbf{w})}$$

Naïve Bayes

Assume, against all common sense, that the words are conditionally independent:

$$
\begin{aligned}
P((1,1,0,0,0,1)|S) &= P(w_1 = 1|S)P(w_2 = 1|S)P(w_3 = 0|S)\times \\
&\quad P(w_4 = 0|S)P(w_5 = 0|S)P(w_6 = 1|S)
\end{aligned}
$$

Naïve Bayes

Assume, against all common sense, that the words are independent:

$$
\begin{aligned}
P[(1,1,0,0,0,1)] \quad &= \quad P(w_1 = 1)P(w_2 = 1)P(w_3 = 0)P(w_4 = 0)\times \\
&\quad\;\; P(w_5 = 0)P(w_6 = 1)
\end{aligned}
$$

Naïve Bayes

$$P(S|\mathbf{w}) = \frac{P(S) \prod_i P(w_i|S)}{\prod_i P(w_i)}$$