CSC-343 ARTIFICIAL INTELLIGENCE
LECTURE 5.3
PROGRAMMING ASSIGNMENT 4

DATA:

emails = [("Where to hide a dead body", 1), ("Grateful dead body of work", 0)]

MODEL: *(Naïve Bayes)*

$$P(\text{Spam}|word1, word2,...wordn) = \alpha \, P(\text{Spam}) \prod P(wordi|\text{Spam})$$

CLASSIFIER:

$$P(\text{spam}| word1, word2,...wordn) > P(\neg \text{spam}| word1, word2,... wordn)$$

B-BOW:

email	Where	To	Hide	A	Dead	Body	Grateful	Of	Work	Spam
1	1	1	1	1	1	1	0	0	0	True
2	0	0	0	0	1	1	1	1	1	False

PART 1 - PRIORS

$$P(\text{Spam}|word1, word2,...wordn) = \alpha \, \underline{P(\text{Spam})} \prod P(wordi|\text{Spam})$$

$P(spam) = count(spam\ emails) / count(all\ emails)$
$P(\neg spam) = count(ham\ emails) / count(all\ emails)$

Spam	P(Spam)
True	p1
False	p2

PART 2 – JOINT PROBABILITIES

$$P(\text{Spam}|word1, word2,...wordn) = \alpha\ P(\text{Spam}) \prod P(wordi|\text{Spam})$$

$$P(wordi\ |\ \text{Spam}) = P(wordi,\ \text{Spam})\ /\ P(\text{Spam})$$

Joint Probabilities: P(wordi, Spam)

jp1 = count(spam emails containing $word_i$)   / count(all emails)
jp2 = count(spam emails NOT containing $word_i$) / count(all emails)
jp3 = count(ham emails containing $word_i$)   / count(all emails)
jp4 = count(ham emails NOT containing $word_i$) / count(all emails)

$word_i$	Spam	P(word$_i$, Spam)		$word_i$	Spam	P(word$_i$, Spam)		$word_i$	Spam	P(word$_i$, Spam)
T	T	jp1		T	T	jp1		T	T	jp1
F	T	jp2		F	T	jp2		F	T	jp2
T	F	jp3		T	F	jp3		T	F	jp3
F	F	jp4		F	F	jp4		F	F	jp4

$word_i$	Spam	P(word$_i$, Spam)		$word_i$	Spam	P(word$_i$, Spam)		$word_i$	Spam	P(word$_i$, Spam)
T	T	jp1		T	T	jp1		T	T	jp1
F	T	jp2		F	T	jp2		F	T	jp2
T	F	jp3		T	F	jp3		T	F	jp3
F	F	jp4		F	F	jp4		F	F	jp4

$word_i$	Spam	P(word$_i$, Spam)		$word_i$	Spam	P(word$_i$, Spam)		$word_i$	Spam	P(word$_i$, Spam)
T	T	jp1		T	T	jp1		T	T	jp1
F	T	jp2		F	T	jp2		F	T	jp2
T	F	jp3		T	F	jp3		T	F	jp3
F	F	jp4		F	F	jp4		F	F	jp4

PART 3 – CONDITIONAL PROBABILITIES

$$P(\text{Spam}|word1, word2,...wordn) = \alpha \ P(\text{Spam}) \prod P(word_i|\text{Spam})$$

$$P(word \ i| \text{Spam}) = P(\text{word } i, \text{Spam})/P(\text{Spam})$$

word$_i$	Spam	P(word$_i$| Spam)
T	T	cp1
F	T	cp2
T	F	cp3
F	F	cp4

word$_i$	Spam	P(word$_i$| Spam)
T	T	cp1
F	T	cp2
T	F	cp3
F	F	cp4

word$_i$	Spam	P(word$_i$| Spam)
T	T	cp1
F	T	cp2
T	F	cp3
F	F	cp4

word$_i$	Spam	P(word$_i$| Spam)
T	T	cp1
F	T	cp2
T	F	cp3
F	F	cp4

word$_i$	Spam	P(word$_i$, Spam)
T	T	cp1
F	T	cp2
T	F	cp3
F	F	cp4

word$_i$	Spam	P(word$_i$, Spam)
T	T	cp1
F	T	cp2
T	F	cp3
F	F	cp4

word$_i$	Spam	P(word$_i$| Spam)
T	T	cp1
F	T	cp2
T	F	cp3
F	F	cp4

word$_i$	Spam	P(word$_i$| Spam)
T	T	cp1
F	T	cp2
T	F	cp3
F	F	cp4

word$_i$	Spam	P(word$_i$| Spam)
T	T	cp1
F	T	cp2
T	F	cp3
F	F	cp4

PART 4 - CLASSIFICATION

email	P(spam | email)	P(¬spam | email)	Our Prediction	Data
	$= P(\text{spam}|word_1, word_2,...word_n)$ $= P(\text{spam}) \prod P(wordi|\text{spam})$ $= P(\text{spam}) \ cp1$	$= P(\text{¬spam}|word1, word2,...wordn)$ $= P(\text{¬spam}) \prod P(wordi|\text{¬spam})$ $= P(\text{¬spam}) \ cp3$	P(spam | email) > P(¬spam | email)	Spam
1				True
2				False

PART 5 – ACCURACY

Accuracy = Count(Our Prediction == Prediction from Data) / Count(Emails)