Problem 1 Page 1

Sentence:

1. good film, recommend ~~t~~ highly	+
2. good acting, good direction, and fun	+
3. good, good, good. fun, fun, fun	+
4. not good, highly avoidable	−
5. not fun, boring	−

Test data

1. Boring, not good
2. Do not recommend ~~t~~ (Note: I split don't into "do not")

Standard naive Bayes:

$P(-) = \frac{2}{5}$ $\quad P(+) = \frac{3}{5}$ $\quad V = 11$

Test sentence 1: ~~Positive~~ Negative

$P(\text{Boring}|+) = \frac{1}{26}$ $\quad P(\text{Boring}|-) = \frac{1+1}{7+11} = \frac{2}{18}$

$P(\text{not}|+) = \frac{1}{26}$ $\quad P(\text{not}|-) = \frac{2+1}{7+11} = \frac{3}{18}$

$P(\text{good}|+) = \frac{6+1}{15+11} = \frac{7}{26}$ $\quad P(\text{good}|-) = \frac{1+1}{7+11} = \frac{2}{18}$

~~$P(+)P(S|+) = \frac{3}{5} \cdot \frac{7}{26} = \frac{7}{25}$~~ ~~[illegible]~~

$P(+)P(S|+) = \frac{3}{5} \cdot \frac{1}{26} \cdot \frac{1}{26} \cdot \frac{7}{26} = \frac{21}{87{,}880}$ $\quad P(-)P(S|-) = \frac{2}{5} \cdot \frac{2}{18} \cdot \frac{3}{18} \cdot \frac{2}{18} = \frac{1}{1215}$

Label: ~~Positive~~ Negative

~~Because 2/3 words in the positive calculation were unknown, it skewed the calculation. Not ignoring unknown words might help as well (and using the Laplace smoothing to prevent 0 values).~~

Problem 2 page 2

Test Sentence 2: Negative

$P(Do|+) = unk$
$P(not|+) = \frac{1}{26}$
$P(recommend|+) = \frac{1+1}{15+11} = \frac{2}{26}$

$P(Do|-) = unk$
$P(not|-) = \frac{2+1}{7+11} = \frac{3}{18}$
$P(recommend|-) = \frac{1}{18}$

$P(+)P(S|+) = \frac{3}{5} \cdot \frac{1}{26} \cdot \frac{2}{26} = \frac{3}{1690}$

$P(-)P(S|-) = \frac{2}{5} \cdot \frac{3}{18} \cdot \frac{1}{18} = \frac{1}{270}$

Label: Negative

Binary Naive Bayes: $V = 11$

Test Sentence 1: Negative

$P(Boring|+) = \frac{1}{20}$
$P(not|+) = \frac{1}{20}$
$P(good|+) = \frac{3+1}{9+11} = \frac{4}{20}$

$P(Boring|-) = \frac{2}{18}$
$P(not|-) = \frac{3}{18}$
$P(good|-) = \frac{2}{18}$

$P(+)P(S|+) = \frac{3}{5} \cdot \frac{1}{20} \cdot \frac{1}{20} \cdot \frac{4}{20} = \frac{1}{5000}$

$P(-)P(S|-) = \frac{3}{5} \cdot \frac{2}{18} \cdot \frac{2}{18} \cdot \frac{2}{18} = \frac{1}{1215}$

Label: Negative

Test Sentence 2: Negative

$P(Do|+) = unk$
$P(not|+) = \frac{1}{20}$
$P(recommend|+) = \frac{1+1}{9+11} = \frac{2}{20}$

$P(Do|-) = unk$
$P(not|-) = \frac{2+1}{7+11} = \frac{3}{18}$
$P(recommend) = \frac{1}{18}$

$P(+)P(S|+) = \frac{3}{5} \cdot \frac{1}{20} \cdot \frac{2}{20} = \frac{3}{1000}$

$P(-)P(S|-) = \frac{3}{5} \cdot \frac{3}{18} \cdot \frac{1}{18} = \frac{1}{270}$

Label: Negative

Problem 2 page 2

Test Sentence 2: Negative

$P(Do|+) = unk$
$P(not|+) = \frac{1}{26}$
$P(recommend|+) = \frac{1+1}{15+11} = \frac{2}{26}$

$P(Do|-) = unk$
$P(not|-) = \frac{2+1}{7+11} = \frac{3}{18}$
$P(recommend|-) = \frac{1}{18}$

$P(+)P(S|+) = \frac{3}{5} \cdot \frac{1}{26} \cdot \frac{2}{26} = \frac{3}{1690}$

$P(-)P(S|-) = \frac{2}{5} \cdot \frac{3}{18} \cdot \frac{1}{18} = \frac{1}{270}$

Label: Negative

Binary Naive Bayes: $V = 11$

Test Sentence 1: Negative

$P(Boring|+) = \frac{1}{20}$
$P(not|+) = \frac{1}{20}$
$P(good|+) = \frac{3+1}{9+11} = \frac{4}{20}$

$P(Boring|-) = \frac{2}{18}$
$P(not|-) = \frac{2}{18}$
$P(good|-) = \frac{2}{18}$

$P(+)P(S|+) = \frac{3}{5} \cdot \frac{1}{20} \cdot \frac{1}{20} \cdot \frac{4}{20} = \frac{1}{5000}$

$P(-)P(S|-) = \frac{3}{5} \cdot \frac{2}{18} \cdot \frac{2}{18} \cdot \frac{2}{18} = \frac{1}{1215}$

Label: Negative

Test Sentence 2: Negative

$P(Do|+) = unk$
$P(not|+) = \frac{1}{20}$
$P(recommend|+) = \frac{1+1}{9+11} = \frac{2}{20}$

$P(Do|-) = unk$
$P(not|-) = \frac{2+1}{7+11} = \frac{3}{18}$
$P(recommend) = \frac{1}{18}$

$P(+)P(S|+) = \frac{3}{5} \cdot \frac{1}{20} \cdot \frac{2}{20} = \frac{3}{1000}$

$P(-)P(S|-) = \frac{2}{5} \cdot \frac{3}{18} \cdot \frac{1}{18} = \frac{1}{270}$

Label: Negative