

## Big Data Processing

Shubham Gupta  
(202318052)

### Question 1

	cat	Documents
Training	-	just plain boring
	-	entirely predictable and lots energy
	-	no surprises and very good laughs
	+	very powerful
	+	the most for film of the summer
Testing	?	predictable with no fun

ANSWER:  $P(w/c) = \frac{\text{count}(w,c) + 1}{\text{count}(c) + |V|}$  ;  $P(-) = \frac{3}{5}$  ;  $P(+) = \frac{2}{5}$  here  $|V| = 20$

$$P(\text{predictable} / -) = \frac{1+1}{13+20} = \frac{2}{33} ; P(\text{with} / -) = \frac{0+1}{13+20} = \frac{1}{33}$$

$$P(\text{predictable} / +) = \frac{0+1}{8+20} = \frac{1}{28} ; P(\text{with} / +) = \frac{0+1}{8+20} = \frac{1}{28}$$

$$P(\text{no} / -) = \frac{1+1}{13+20} = \frac{2}{33} ; P(\text{fun} / -) = \frac{0+1}{13+20} = \frac{1}{33}$$

$$P(\text{no} / +) = \frac{0+1}{8+20} = \frac{1}{28} ; P(\text{fun} / +) = \frac{1+1}{8+20} = \frac{2}{28} = \frac{1}{14}$$

$$P(- / \text{Test}) = \frac{3}{5} \times \frac{2}{33} \times \frac{1}{33} \times \frac{2}{33} \times \frac{1}{33} = \frac{12}{5929605}$$

$$P(+ / \text{Test}) = \frac{2}{5} \times \frac{1}{28} \times \frac{1}{28} \times \frac{1}{28} \times \frac{1}{14} = \frac{2}{1536640}$$

$\therefore P(- / \text{Test}) > P(+ / \text{Test})$  ; Test data belongs to - class