

Name : ~~Barot~~ Barot Kushal

BIG DATA

ID : 202318006

	Cat	Document
training	-	Just plain boring
	-	entirely predictable and lacks energy
	-	no surprises and very few laughs
	+	very powerful
	+	the most fun film of the summer
test	?	predictable with no fun

① Prior from training:

$$P(c_j) = \frac{N_{c_j}}{N_{\text{total}}} \Rightarrow \begin{aligned} P(-) &= 3/5 \\ P(+) &= 2/5 \end{aligned}$$

② Likelihood from training:

$$P(w_i | c) = \frac{\text{count}(w_i, c) + 1}{(\sum_{w \in V} \text{count}(w, c)) + |V|}$$

Here  $|V| = 20$  (i.e. unique words)

$$\therefore P(\text{predictable} | +) = \frac{0+1}{1+20} = \frac{1}{21}$$

$$P(\text{Predictable} | -) = \frac{1+1}{14+20} = \frac{2}{34} = \frac{1}{17}$$



$$P(\text{no} | +) = \frac{0+1}{9+20} = \frac{1}{29}$$

$$P(\text{no} | -) = \frac{1+1}{14+20} = \frac{1}{17}$$

$$P(\text{fun} | +) = \frac{1+1}{9+20} = \frac{1}{29}$$

$$P(\text{fun} | -) = \frac{0+1}{14+20} = \frac{1}{34}$$

Here we have drop 'with' because 'with' is not present in the training cat.

③ Solving test set :

$$P(-) \cdot P(\text{Pred} | -) \cdot P(\text{No} | -) \cdot P(\text{Fun} | -)$$

$$= \frac{3}{5} \times \frac{2}{34} \times \frac{2}{34} \times \frac{1}{34} = 6.105 \times 10^{-5}$$

$$P(+). P(\text{Pred} | +) \cdot P(\text{No} | +) \cdot P(\text{fun} | +)$$

$$= \frac{2}{5} \times \frac{1}{29} \times \frac{1}{29} \times \frac{2}{29}$$

$$= \underline{\underline{3.2 \times 10^{-5}}}$$