

Name: Satyam Marwaha

Big Data

ID: 202318626

cat Document

- train
- Just Plain boring
 - entirely Predictable and lacks energy
 - no surprises and very few laughs
 - + very powerful
 - + the most fun film of the summer.

Test 1 Predictable with 10 fun

① Prior from training

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

② Likelihood from training

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

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

$$\therefore P(\text{Predictable} | +) = \frac{0 + 1}{9 + 20} = \frac{1}{29}$$

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

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

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

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

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

here we have prob 'with' because 'with' is not present in the friends set.

③ solving test set :

$$\begin{aligned} P(-) \cdot P(\text{pref} | -) \cdot P(\text{no} | -) \cdot P(\text{fun} | -) \\ = \frac{3}{5} \cdot \frac{2}{34} \times \frac{2}{34} \times \frac{1}{34} \\ = 6.163 \times 10^{-5} \end{aligned}$$

$$\begin{aligned} P(+) \cdot P(\text{pref} | +) \cdot P(\text{no} | +) \cdot P(\text{fun} | +) \\ = \frac{2}{5} \times \frac{1}{29} \times \frac{1}{29} \times \frac{2}{29} \\ = 3.2 \times 10^{-5} \end{aligned}$$