

Exercise

	Cat	Documents
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

Solution:

Collection of vocabulary: just, plain, boring, entirely, predictable, and lacks, energy, no, surprises, very, few, laughs, powerful the, most, fun, film, of, summer

• Prior from training

$$P(C_j) = \frac{N_{C_j}}{N_{\text{total}}}$$

$$\Rightarrow P(-) = 3/5 \quad \& \quad P(+) = 2/5$$

• likelihood from training

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

Given: $|V|$ = no. of unique words = 20

$$\therefore P(\text{Predictable}/+) = \frac{1}{29} \quad \left(\because \frac{0+1}{9+20} \right)$$

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

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

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

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

$$P(\text{fun}/+) = \frac{1+1}{9+20} = \frac{2}{29}$$

• Hence, In test set :

$$\therefore P(-) \cdot P(\text{pred}|-) \cdot P(\text{No}|-) \cdot P(\text{fun}|-)$$

$$= \frac{3}{5} \times \frac{2}{34} \times \frac{2}{34} \times \frac{1}{34} = 6.106 \times 10^{-5} \quad \text{--- ①}$$

$$\therefore P(+). P(\text{pred}|+). P(\text{No}|+). P(\text{fun}|+)$$

$$= \frac{2}{5} \times \frac{1}{29} \times \frac{1}{29} \times \frac{2}{29} = 3.2 \times 10^{-5} \quad \text{--- ②}$$

~~As, ① > ②~~

So, we can say that the test document is
of negative category.