

Week 3 Group Exercise Solutions

Text Cat/NB/Sentiment

Part 1: Group Exercise

Test sentence: predictable with no originality

Since “originality” and “with” don’t appear in the vocabulary from the training set, we can ignore them. So the new set of words becomes: “predictable,” “no”

$$|V| = 20$$

Priors:

$$P(+) = \frac{2}{5}$$

$$P(-) = \frac{3}{5}$$

Conditional likelihoods:

$$P(\text{“predictable”} \mid +) = \frac{0+1}{9+20} = \frac{1}{29}$$

$$P(\text{“predictable”} \mid -) = \frac{1+1}{14+20} = \frac{2}{34}$$

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

$$P(\text{“no”} \mid -) = \frac{1+1}{14+20} = \frac{2}{34}$$

$$P(+ \mid \text{“predictable”, “no”}) = \frac{1 * 1 * 2}{29 * 29 * 5}$$

$$P(- \mid \text{“predictable”, “no”}) = \frac{2 * 2 * 3}{34 * 34 * 5}$$

$P(- \mid \text{“predictable”, “no”}) > P(+ \mid \text{“predictable”, “no”})$, so predict **negative**.

Without add-1 smoothing, the conditional probability of the positive class would be 0.

Part 2: Challenge Problems

“Only boring people would dislike this movie.”

“Nobody likes this movie.”

“Not an Oscar winner.”

“... mafia , rap stars and hood rats butt their ugly heads in a regurgitation of cinematic violence that gives brutal birth to an unlikely , but likable , hero “

Binary NB may work better on sentences that have a lot of unnecessarily repeated words.

Also, when a single word or a few words have huge importance on the classification, we may want these to hold more importance and remove the noise of other more common words, in which case we may choose to use binary NB instead of full count NB