

9 Machine Learning and Real-world Data (av308)

An online brand-safety company has as its goal to warn its clients when there is a problem with their brand name on social media. To this end, it wants to develop a classifier to determine whether a brand is being attacked on social-media platforms. To develop this classifier, the company collected 5000 social-media posts: 100 posts referring to each of 50 brands. For each brand, company employees read all 100 posts, and hand-annotated the brand with one of the two labels, *attacked* and *safe*; this resulted in 10 brands being labelled *attacked* and 40 brands labelled *safe*. Your task is to develop a naive Bayes classifier that uses the text of the posts as features.

- (a) Give the equations for a naive Bayes classifier for the task of determining whether a brand is safe or being attacked. [2 marks]
- (b) How would you split the data into training and testing? Justify your choice. [2 marks]
- (c) Here are some posts for two brands used for training:

BRAND	LABEL	POSTS
GStuff	safe	Awesome products as always from GStuff!
GStuff	safe	Fantastic customer support from GStuff!
GStuff	safe	Awesome GStuff product, I wish it were cheaper
AThing	attacked	Not the best experience from AThing
AThing	attacked	Will not buy AThing again! Terrible performance
AThing	attacked	Terrible performance from AThing, avoid

- (i) What features do you expect your naive Bayes classifier to consider important? Give three examples for the 10 *attacked* brands and also three examples for the 40 *safe* brands. How well do you expect each of these features to generalize? [6 marks]
- (ii) Based on your observations in Part (c)(i), suggest and justify two changes to feature extraction to improve generalization. [4 marks]
- (d) You claim that a future product will enable you to warn clients about attacks before any competitor does. Give three modifications to your classifier and the evaluation setup that might help you achieve this. [3 marks]
- (e) Consider changing your approach to a classifier operating at the post level, i.e. classifying media posts instead of brands. There are advantages and disadvantages of doing so. Give three. [3 marks]