

# CS 725: Foundations of Machine Learning

## Homework 3

Soumen Kumar Mondal  
23m2157@iitb.ac.in

Naay Balodia  
23m2166@iitb.ac.in

September 17, 2023

### Abstract

In this assignment, we will implement Naive Bayes model on a toy datasets for 3 class classification task. Our key goal in this assignment is to correctly implement these models and analyze the results we obtained.

## 1 Naive Bayes Classifier

In the classification problem, the features ( $\bar{x}$ ) will be given and we are interested to find out what is the probability that a label ( $\hat{y}$ ) will be correctly classified? More generally, given the features, what is the most likely label?

In this homework, it is given that  $\bar{x} = [x_1, x_2, \dots, x_{10}]^T$  that means there are 10 features in the dataset. The class label  $y$  can be any one of the class between  $[0, 1, 2]$ . Probability of the label  $y$  given the feature  $\bar{x}$  is denoted as  $P[y|\bar{x}]$ . In this homework problem, since there are 3 classes, we will be interested to know the vector of posterior probabilities of labels for all the classes as:

$$P[y = 0|\bar{x}], P[y = 1|\bar{x}], \dots, P[y = 2|\bar{x}] \quad (1)$$

The output of the classifier will be given as:

$$\hat{y} = \operatorname{argmax}_{k=0,1,2} P[y = k|\bar{x}] \quad (2)$$

The posterior probability of label term can be written in terms of prior probability of label and likelihood of features given a label by Bayes theorem as follows:

$$\text{for } k = 0, 1, 2 : \quad P[y = k|\bar{x}] = \frac{P[\bar{x}|y = k]}{P[\bar{x}]} \quad (\text{Bayes Theorem}) \quad (3)$$

The denominator term  $P[\bar{x}]$  doesn't depend on the label  $y$ . Therefore, we can even ignore this term as we are more interested in the likelihood or a score rather than the valid probability measure. In the numerator, the likelihood term can be expanded for all the features that are conditionally dependent on other features and label by multiplicative rule of conditional probability as follows:

$$P[\bar{x}|y = k] = P[x_1 \cap x_2 \cap \dots \cap x_{10}|y = k] \quad (4)$$

$$= P[x_1|y = k] \cdot P[x_2|x_1 \cap y = k] \dots P[x_{10}|x_1 \cap x_2 \cap \dots \cap x_9 \cap y = k] \quad (5)$$

In general, the features are not independent of each other but given a label, it is naively assumed that the features become conditionally independent under the given label. Hence the above equation can be simplified as follows:

$$P[\bar{x}|y = k] = P[x_1|y = k] \cdot P[x_2|y = k] \dots P[x_{10}|y = k] \quad (6)$$

$$= \prod_{i=1}^{10} P[x_i|y = k] \quad (7)$$

$$P[y = k|\bar{x}] = \left( \prod_{i=1}^{10} P[x_i|y = k] \right) \cdot P[y = k] \quad (8)$$

$$\hat{y} = \operatorname{argmax}_{k=0,1,2} \left( \prod_{i=1}^{10} P[x_i|y = k] \right) \cdot P[y = k] \quad (9)$$