# 1 introduction

*Naive Bayes* is a simple technique for constructing classifiers, which assumes that the value of a particular feature is independent of the value of any other feature, thus its naivity.

# 2 probabilistic model

using Bayesian probability, ïnaiveässumption we have

$$p(C_k|x_1, ..., x_n) = \frac{p(Y = C_k)\prod_j p(X^{(j)} = x^{(j)}|Y = c_k)}{\sum_k p(Y = C_k)\prod_j p(X^{(j)} = x^{(j)}|Y = c_k)}, \tag{1}$$

so we have our naive Bayesian classifier:

$$y = f(x) = argmax_{c_k}\frac{p(Y = C_k)\prod_j p(X^{(j)} = x^{(j)}|Y = c_k)}{\sum_k p(Y = C_k)\prod_j p(X^{(j)} = x^{(j)}|Y = c_k)}, \tag{2}$$

notice that the $\sum_k p(Y = C_k)\prod_j p(X^{(j)} = x^{(j)}|Y = c_k)$ is same for all $c_k$, we have:

$$y = argmax_{c_k} p(Y = C_k)\prod_j p(X^{(j)} = x^{(j)}|Y = c_k). \tag{3}$$

we choose our loss function as:

$$L(Y, f(X)) = \begin{cases} 1, & Y \neq f(X) \\ 0, & Y = f(X) \end{cases} \tag{4}$$

thus, from our goal to minimize the expectation risk, we get our principle to maximize posterior probability.

# 3 parameter estimation

a simple estimation is:

$$p(Y = c_k) = \frac{\sum_{i=1}^N I(y_i = c_k)}{N},\ k = 1, ..., K, \tag{5}$$

$$p(X^{(j)} = a_{jl}|Y = c_k) = \frac{\sum_{i=1}^N I(x_i^{(j)} = a_{jl},\ y_i = c_k)}{\sum_{i=1}^N I(y_i = c_k)} \tag{6}$$

but we see, sometimes the probability comes zero just because we have not seen this class in our training set, so another method (*Laplace smoothing*) is:

$$p(Y = c_k) = \frac{\sum_{i=1}^N I(y_i = c_k) + \lambda}{N + \lambda} \tag{7}$$

$$p(X^{(j)} = a_{jl}|Y = c_k) = \frac{\sum_{i=1}^N I(x_i^{(j)} = a_{jl},\ y_i = c_k) + \lambda}{\sum_{i=1}^N I(y_i = c_k) + S_j\lambda}. \tag{8}$$