A Note on an Elementary Formulation of the Naive Bayesian Model

jiyucho9145

February 9, 2019

Abstract

Some mail server programs and some mail client programs classify messages into several categories by the naive Bayesian classifier automatically. The algorithm of the naive Bayesian classifier is based on the naive Bayesian model, which satisfies the two specific assumptions.

In this note, we define a probabilistic model (message receiving model) without the two assumptions and show that the message receiving model coincides with the naive Bayesian model under the two assumptions.

Contents

1 Introduction

The naive Bayesian classifier is used in some mail server programs and some mail client programs. Many Japanese books regarding to machine leaning have been published in recent years. Several popular machine larning algorithms are introduced in these books. The naive Bayesian classifier is one of these algorithms.

The mathematical background of the naive Bayesian classifier is not introduced in some books. This policy is appropriate in practice because programmers are able to implement and use the naive Bayesian classifier without the mathematical background.

However, we have an interest in the mathematical background of the naive Bayesian classifier. We formulate the naive Bayesian model in elementary way by probability theory because the naive Bayesian classifier is based on the naive Bayesian model.

Next section introduces a thory on a probabilistic model (message receiving model), which is a base of the theory on the naive Bayesian model, section 3 formulates the naive Bayesian model.

2 Message Receiving Model

2.1 Message Receiving Model

Definition 2.1. Let W be an nonempty finite set, $E = \mathrm{Map}(W, \{0, 1\}), C = \{c_1, c_2\}, c_1 \neq c_2$, then we call (W, E, C) a message receiving model in this note.

Definition 2.2. Let $g : F \to C, F \subset E, F \neq \emptyset, F_{g,i} = g^{-1}(c_i) \neq \emptyset$, then we call $g : F \to C$ a training data for a message receiving model (W, E, C) in this note.

Definition 2.3. Let $e \in E$, then we define a support $\mathrm{Supp}(e)$ of e as follows:

$$\mathrm{Supp}(e) = \{w \in W; e(w) \neq 0\}. \tag{1}$$

Definition 2.4. Let $Z \subset W$, then we define a set $M_g(Z)$ as follows:

$$M_g(Z) = \{e \in F; Z \subset \mathrm{Supp}(e)\}. \tag{2}$$

Definition 2.5. Let $w \in W$, then we define a set $M_g(w)$ as follows:

$$M_g(w) = M_g(\{w\}). \tag{3}$$

Definition 2.6. Let $e \in E$, then we define a set $M_g(e)$ as follows:

$$M_g(e) = \cap_{w \in \mathrm{Supp}(e)} M_g(w). \tag{4}$$

2.2 Message Receiving Measures

Definition 2.7. We call the map
$$m_g : \mathrm{Pow}(F) \to \mathbb{R}; G \mapsto \mathrm{card}(G) \tag{5}$$

a message receiving measure of $g : F \to C$ in this note. Where $\mathrm{Pow}(S)$ denots the power set of S for arbitrary set S.

Proposition 2.1. $m_g : \mathrm{Pow}(F) \to \mathbb{R}$ is a measure on measurable space $(F, \mathrm{Pow}(F))$.

Proposition 2.2.
$$\mathrm{card}(F_{g,1}) + \mathrm{card}(F_{g,2}) = \mathrm{card}(F). \tag{6}$$

2.3 Message Receiving Probabilities

Definition 2.8. We call the map

$$P_g : \mathrm{Pow}(F) \to \mathbb{R}; G \mapsto m(G)/m(F) \tag{7}$$

a message receiving probability of $g : F \to C$ in this note.

Proposition 2.3. $P_g : \mathrm{Pow}(F) \to \mathbb{R}$ is a probability (measure) on measurable space $(F, \mathrm{Pow}(F))$.

Proposition 2.4.

$$P_g(F_{g,1}) + P_g(F_{g,2}) = 1. \tag{8}$$

2.4 Message Receiving Conditional Probabilities

Definition 2.9. We call the conditional probability

$$P_g(-|-) : \mathrm{Pow}(F) \times \mathrm{Pow}(F) \to \mathbb{R} \tag{9}$$

a message receiving conditional probability of $g : F \to C$ in this note.

Proposition 2.5. For arbitrary $M, N \subset F$, following equation is satisfied:

$$P_g(M \cap N) = P_g(M|N)P_g(N). \tag{10}$$

Proposition 2.6. For arbitrary $M \subset F$, following equation is satisfied:

$$P_g(F_{g,1}|M) = \frac{P_g(M|F_{g,1})P(F_{g,1})}{P_g(M|F_{g,1})P(F_{g,1}) + P_g(M|F_{g,2})P(F_{g,2})}. \tag{11}$$

Proposition 2.7. For arbitrary $M \subset F$, following equation is satisfied:

$$P_g(F_{g,1}|M) + P_g(F_{g,2}|M) = 1. \tag{12}$$

3 Comparation with Naive Bayesian Model

3.1 Calculationg Conditional Probabilities under the Two Specific Assumptions

Theorem 3.1. Assume that $g : F \to C$ satisfies following two conditions: (i) $P_g(F_{g,1}) = P_g(F_{g,2}) = 1/2$; (ii) for arbitrary $w_1, w_2, \ldots, w_r \in W$,

$$P_g(M_g(\{w_1, w_2, \ldots, w_r\})|F_{g,i}) = P_g(M_g(w_1)|F_{g,i})P_g(M_g(w_2)|F_{g,i}) \cdots P_g(M_g(w_r)|F_{g,i}). \tag{13}$$

i.e, $M_g(w_1), M_g(w_2), \ldots, M_g(w_r)$ are independent. And, let $Q_g(w_1, w_2, \ldots, w_r), R_g(w_1, w_2, \ldots, w_r)$ be following functions:

$$Q_g(w_1, w_2, \ldots, w_r) = \prod_i^r P_g(M(w_i)|F_{g,1}), \tag{14}$$

$$R_g(w_1, w_2, \ldots, w_r) = \prod_i^r \left(\frac{P_g(M(w_i)) - P_g(M(w_i)|F_{g,1})P_g(F_{g,1})}{1 - P_g(F_{g,1})} \right). \tag{15}$$

Then, for arbitrary $e \in E$, the following equation is satisfied

$$P_g(F_{g,1}|M_g(e)) = \frac{Q_g(w_1, w_2, \ldots, w_r)}{Q_g(w_1, w_2, \ldots, w_r) + R_g(w_1, w_2, \ldots, w_r)}. \tag{16}$$

Where $\mathrm{Supp}(e) = \{w_1, w_2, \ldots, w_r\} \subset W$.