

===== Bozeman, MT 59715, USA
 ~~~~~~ d7dc344301a47614a62fc3e9a618b5de03f3715b

## Unnamed Project 1 Writeup

**Ethan Fison**

ETHANFISON@GMAIL.COM

**Zan Monteith**

EMAIL

**Alex Salois**

EMAIL

**Sage Acteson**

EMAIL

*Gianforte School of Computing*

*Montana State University*

*~~~~~ HEAD Bozeman, MT 59715, USA*

**Editor:** Names

### Abstract

This paper describes the Naïve Bayes algorithm, used for classification of data. This algorithm builds its model by finding the average value for each attribute of a given class, then classifies an input by finding the class it most closely matches. In this experiment, we run ten-fold cross-validation of our models build from 5 different datasets acquired from (whatever that repository is) to test the accuracy of the algorithm

**Keywords:** Naïve Bayes, Classification,

### 1. Introduction

Probabilistic inference has become a core technology in AI, largely due to developments in graph-theoretic methods for the representation and manipulation of complex probability distributions (?). Whether in their guise as directed graphs (Bayesian networks) or as undirected graphs (Markov random fields), *probabilistic graphical models* have a number of virtues as representations of uncertainty and as inference engines. Graphical models allow a separation between qualitative, structural aspects of uncertain knowledge and the quantitative, parametric aspects of uncertainty...

*Remainder omitted in this sample. See <http://www.jmlr.org/papers/> for full paper.*

### Acknowledgments

We would like to acknowledge support for this project from the National Science Foundation (NSF grant IIS-9988642) and the Multidisciplinary Research Program of the Department of Defense (MURI N00014-00-1-0637).

## Appendix A.

In this appendix we prove the following theorem from Section 6.2:

**Theorem** *Let  $u, v, w$  be discrete variables such that  $v, w$  do not co-occur with  $u$  (i.e.,  $u \neq 0 \Rightarrow v = w = 0$  in a given dataset  $\mathcal{D}$ ). Let  $N_{v0}, N_{w0}$  be the number of data points for which  $v = 0, w = 0$  respectively, and let  $I_{uv}, I_{uw}$  be the respective empirical mutual information values based on the sample  $\mathcal{D}$ . Then*

$$N_{v0} > N_{w0} \Rightarrow I_{uv} \leq I_{uw}$$

*with equality only if  $u$  is identically 0.* ■

**Proof.** We use the notation:

$$P_v(i) = \frac{N_v^i}{N}, \quad i \neq 0; \quad P_{v0} \equiv P_v(0) = 1 - \sum_{i \neq 0} P_v(i).$$

These values represent the (empirical) probabilities of  $v$  taking value  $i \neq 0$  and 0 respectively. Entropies will be denoted by  $H$ . We aim to show that  $\frac{\partial I_{uv}}{\partial P_{v0}} < 0 \dots$

*Remainder omitted in this sample. See <http://www.jmlr.org/papers/> for full paper.*

## References