TDT4173 Assignment 4

Bayesian Learning & Expectation Maximization

Iver Jordal

# 1. Theory

## Bayesian Learning

### 1.1

My hunch is that likelihood is the same as probability. However, that is not true. “Likelihood is the hypothetical probability that an event that has already occurred would yield a specific outcome. The concept differs from that of a probability in that a probability refers to the occurrence of future events, while a likelihood refers to past events with known outcomes.” [0]

In other words, *likelihood* is the probability of observing data given a hypothesis .

What is the difference between a *maximum a posteriori* hypothesis and a *maximum likelihood* hypothesis?

The maximum a posteriori hypothesis is the single most probable hypothesis given the evidence, i.e. the that maximizes .

The Maximum Likelihood hypothesis is the hypothesis that maximizes .

In the case of a uniform prior, “MAP learning reduces to choosing an that maximizes ” [1]. So in some cases, and are the same hypothesis, but not always. Let me give an example where they are not equal: Let’s say that we’re flipping a coin. The coin is either a normal coin () or one of those special coins that don’t have normal flip probabilities (, ). In an experiment we try find out by flipping the coin several times.

Example of non-uniform prior distribution over :

Uniform prior distribution over :

In the latter case (uniform) but in the former case (non-uniform)

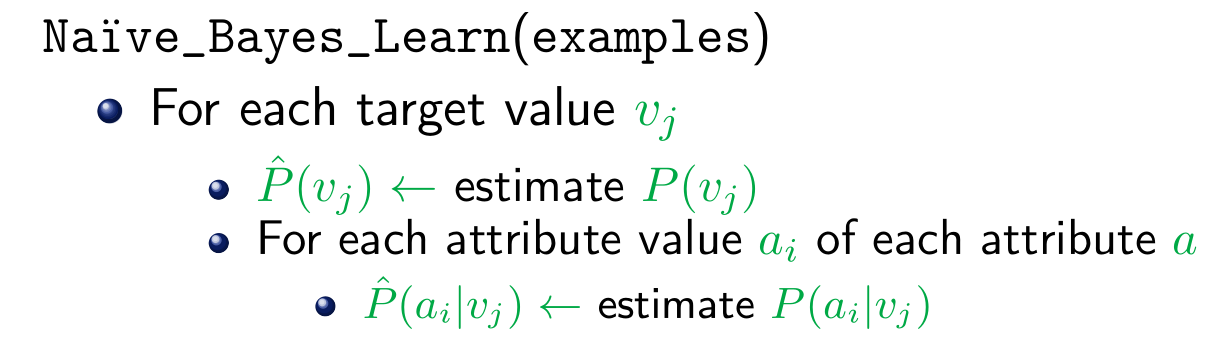
To explain this better, let’s try to express differently:

Because P(D) is constant in this context, we can disregard it. And when we have a uniform prior, is also a constant. So in that case, because

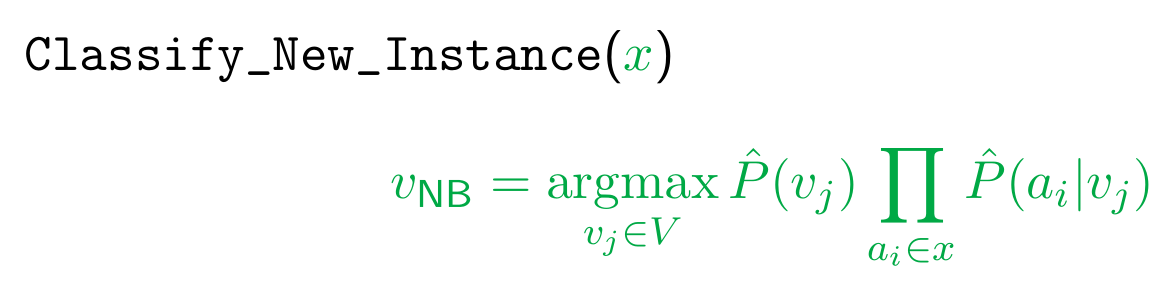
### 1.2

Naive Bayes: You assume that the attributes are conditionally independent of each other

Building a Bayes classifier is all about learning naïve bayes distributions from observations. So how does one train that model? Actually, it’s not that complicated, because it is basically a counting problem. This is the algorithm (copied from a slide in lecture 7):



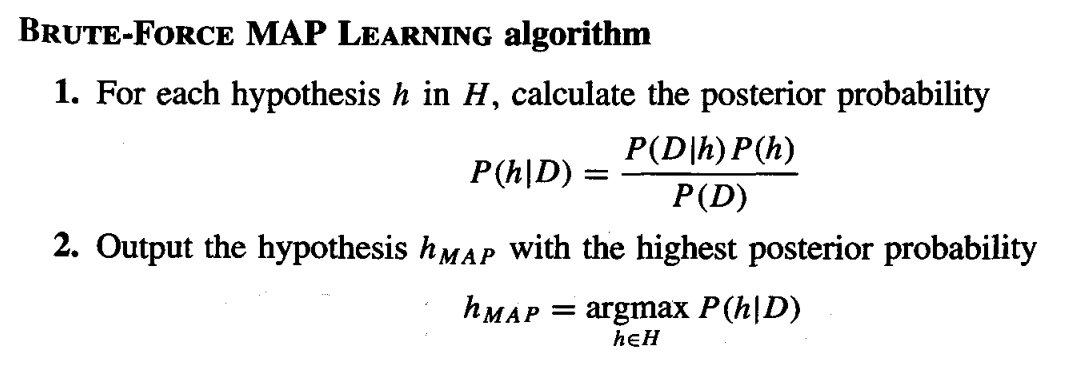
Classifying a new instance with naive bayes:



OptimalBayes classifier:



Is the Brute-force MAP learning algorithm of any use?



Is OptimalBayes classifier an example of a NaiveBayes classifier, or is it the reverse case, or none? My hunch: None

### 1.3

|  |  |  |
| --- | --- | --- |
|  | NaiveBayes | OptimalBayes |
| Computational cost | Moderate. One can use this in practice. | Very high. Almost impossible to use on real-world problems. |
| Performance | Suboptimal | Optimal |
| Considered hypothesis space | A subspace | All of them |

### 1.4

The question is a bit vague. I’m not an expert, so frankly, I don’t have good ideas about how to improve OptimalBayes and NaiveBayes. But anyway, I’ll *try*:

NaiveBayes: Assumes that all variables are conditionally independent, while they are actually not. Performance could probably be improved by adding more hypotheses, somehow. That would probably hurt computational cost, though.

OptimalBayes: Is way too computationally expensive to be used on problems that are not very small. If there would be some way to reduce the considered hypothesis space in a smart way to reduce the computational cost, that would be nice. It would perhaps hurt predictive performance a bit.

### 1.5

What is the difference between NaiveBayes classifier and Bayesian belief network? What is the relation between these two methods?

Bayesian Belief networks describe conditional independence among subsets of variables. In NaiveBayes all variables are conditionally independent.

## 2 Programming

### Expectation Maximization

Ok, so I’m going to implement an EM algorithm, and I have this data set that I’m going to run it on. Cool.

# References

0: Weisstein, Eric W. "Likelihood." From MathWorld--A Wolfram Web Resource. <http://mathworld.wolfram.com/Likelihood.html>

1: Artificial intelligence – a modern approach 3E, Russell & Norvig, Pearson 2010