AI/ML Guild Lab #1
Bayesian Classification
September 22, 2021

1 Introduction

Thomas Bayes: minister, mathematician. Born in London 1701, published "An Essay towards solving a Problem in the Doctrine of Chances" in 1763. It was basically a treatise on card counting.

2 Naive Bayes Classification

Basic Bayesian Statistics

- Conditional Probability: The probability of observing some event, X, given that another event, Y has already been observed. Denoted: $P(X \mid Y)$
- Prior probability: A probability distribution, $P(X)$, that expresses one's belief in an outcome before any evidence is collected.
- Posterior probability: The probability of an event after all the predictor information has been incorporated. Posterior probabilities reflect the uncertainty of assessing an observation to particular class.

$$P(\theta \mid X) = \frac{P(X \mid \theta)P(\theta)}{P(X)}$$

2.1 Bayes Theorem

$$P(\theta \mid X) = \frac{P(X \mid \theta)P(\theta)}{P(X)}$$

This is just the posterior probability, derived in Bayes' original essay.
$P(\theta)$ is the prior distribution
The likelihood $P(X \mid \theta)$ is the evidence of θ provided by X
$P(X)$ is the probability of X given all possible θ.

2.2 Naive Bayes

Exact Bayesian Classification is often impractical, getting the *true* value of the likelihood may be intractable or impossible to acquire. The likelihood is defined as $P(x_1, x_2, \ldots, x_j \mid \theta)$.

We therefore must approximate $P(X \mid \theta)$ by calculating the product of the individual conditional probabilities: $P(x_1 \mid \theta) * P(x_2 \mid \theta) * \cdots * P(x_j \mid \theta)$

$P(X)$ is the same for all values, so we don't even need to calculate it, we can ignore it.

We end up with an approximation of the value we want $P(\theta \mid X)$:

$$P(\theta \mid X) \propto P(X \mid \theta) * P(\theta) = P(\theta) * P(x_1 \mid \theta) * P(x_2 \mid \theta) * \cdots * P(x_j \mid \theta)$$

Classificaiton then becomes assigning the label with the highest likelihood value for a given observation.