COMP4388: **MACHINE LEARNING**

Naive Bayes

Dr. Radi Jarrar

Probabilistic Models

- 80% chance of rain
- 90% chance the tumor is benign
- 88% the email is spam
- 90% chance of winning the elections

- These are probabilistic estimates that show the probability of an event to occur

Probabilistic Models (2)

- Estimates based on probabilistic methods
- Methods concerned with describing uncertainty
- These methods use data on past events to predict future events

Probabilistic Models (3)

- For instance, the weather forecasting of 80% describes that the proportion of days from the past data with similar properties (conditions, attributes) , an 80% chance it will rain which means with similar conditions, 8 times out of 10 times it did rain with similar patters of weather attributes

Probabilistic Models (4)

- 90% chance that the tumor is benign means that based on previously collected data and medical records, analysis of biopsies of similar tissues showed that 9 out of 10 times the tumors were benign (not malignant)

Probabilistic Models (5)

- 88% of previously collected emails showed that based on the pattern of the current email, 8.8 out of 10 times in which emails with similar patters and/or words were spam

Naive Bayes

- Naive Bayes classifier is a generative probabilistic model that based on Bayes Theorem
- It assumes that all features are conditionally independent of one another given the target class
- Naive Bayes classifier also assumes that all assumptions are explicitly built using a set of input feature vectors

Naive Bayes (2)

- Naive Bayes classifier is based on Bayes theorem and it calculates the probability of each target class $y \in Y$ given a feature vector $f \in F$

$$p(y|f) = \frac{p(y)\,p(f|y)}{p(f)}$$

Naive Bayes (3)

$$p(y|f) = \frac{p(y)\,p(f|y)}{p(f)}$$

- where p(f|y) is the unkown probability estimation of the joint distribution of the feature vector *f* and the target class *y; p(y)* is the class prior and is estimated from the training data

Naive Bayes (4)

$$P(spam|Prince) = \frac{P(Prince|spam)P(spam)}{P(Prince)}$$

- P(spam|Prince) is the posterior probability which measures how likely this message spam
- P(Prince|spam) is the likelihood which is the probability that word 'Prince' was used in previous '**spam**' messages
- P(spam) which is the probability that any prior message was spam
- P(Prince) is the marginal likelihood which is the probability that the word 'Prince' appeared in any message

Naive Bayes (5)

- *p(f|y)* is estimated by following Bayes theorem, which assumes that the attributes are independent given the target class

$$p(f_1, f_2, \cdots, f_n \mid y) = p(f_1|y)\ p(f_2|y)\ \ldots\ p(f_n|y)$$

- where f_i is the i^{th} dimension of the feature vector

Naive Bayes (6)

- Though this can be seen as a naive independence assumption, naive Bayes has shown to be a strong classifier in many areas such as Spam email filters
- Moreover, this independence assumption is feasible with the datasets in which the features may not be strongly correlated (as of they are independent)

Naive Bayes (7)

- Naive Bayes classifier can be then written as

$$p(y \mid f) = p(y) \prod_{m=1}^{N} p(f_m \mid y)$$

- and the classification rule of a new input feature vector f_{new} becomes

$$y^* = \arg\max_{y \in S} p(y) \prod_{m=1}^{N} p(f_m^{new} \mid y)$$

Naive Bayes (8)

- Using this classifier, the classification rule becomes as given a new input feature vector, classify it with the target class that of the highest probability

Feature discretisation

- If the classification task involves continuous (quantitative) attributes, they have to be descritised so that the Bayesian classifier can calculate the prior-probability of the target classes
- Algorithms to discretise continuous attributes include Entropy Minimisation Discretisation (EMD), Kernel estimation,…

Strengths of Naive Bayes

- Simple, fast, and quite effective
- Performs well on missing and noisy data
- Does not require many training examples through the training process
- Works well with large number of training examples
- Easy to obtain the estimated probability for a prediction

Weaknesses of Naive Bayes

- Relies on an often faulty assumption of equally important and independent features
- Not the best choice of data with large number of numeric features
- Since the output is estimated probability, target class outputs are more reliable that probabilistic estimations