**PGP in AI/ML**

**Classification - Assignment 2**

The following questions are to enhance your understanding of basic concepts and definitions. You are expected to answer these questions on your own without referring to internet.

1. In a particular pain clinic, 15% of patients are prescribed narcotic pain killers. Overall, six percent of the clinic’s patients are addicted to narcotics (including pain killers and illegal substances). Out of all the people prescribed pain pills, 9% are addicts. If a patient is an addict, what is the probability that they will be prescribed pain pills ? (Hint: Use Bayes Theorem) **[3M**]

**Answer 1.**

Let,

P(NP) : % of people prescribed narcotics pain killer

P(A) : % of people addicted

The given probabilities are,

P(NP) = 0.15 (Prior class probability)

P(A) = 0.06 (predictor prior probability)

P(A|NP) = 0.09 (Likelihood)

Problem to solve is P(NP|A) : % of people who are addict prescribed pain killers

Using Bayes theorem,

P(NP|A) = [ P(A|NP) \* P(NP) ] / P(A)

= 0.09 \* 0.15 / .06

= 0.225

=> Answers is, 0.225

1. Given below is the classification data set of a few mammals/non-mammals, based on a few attributes (Give Birth, Can Fly, Lives In Water, Have Legs).

****

For the input set given below, please determine the Class of the creature using Naïve Bayes classifier **[7M]**

**Give Birth – Yes, Can fly – No, Live in water – Sometimes, Have legs – Yes**

**Answer 2.**

Let,

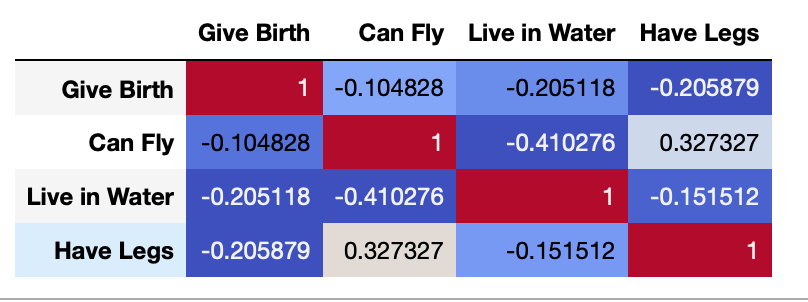
**F1** : Give birth,

**F2** : Can Fly

**F3** : Live in Water

**F4** : Have Legs

P(Class | F1=yes, F2=no, F3=sometimes, F4=yes) be the posterior probability that we need to solve.

As per Bayes classification rule, and assuming conditional independence rule that F1, F2 .. F4 are independent of each other.

This when plotted using correlation matrix shows that the features are fairly independent (corr < 0.5) and Bayes classifier can be applied to this dataset for classification.

So applying Naiive Bayes classification,

P1 *=* P(Class=m| F1=yes, F2=no, F3=sometimes, F4=yes) =

P(class=m) \* P(F1|class=m) \* P(F2|class=m) \* P(F3|class=m) \* P(F4|class=m) / P(F1,F2,F3,F4)

P2 = P(Class=nm| F1=yes, F2=no, F3=sometimes, F4=yes) =

P(class=nm) \* P(F1|class=nm) \* P(F2|class=nm) \* P(F3|class=nm) \* P(F4|class=nm) / P(F1,F2,F3,F4)

And as per classification, C = argmax(p1, p2)

*Note: With the given observation P(f3 = sometime | mammal) doesn’t exist and that could draw the probability to 0 which is not desired. Hence we apply Laplace smoothing for all the probability distributions.*

P(Mammals) = 7 + 1 / 20 + 2\*1

P(Non-Mammals) = 13 + 1 / 20 + 2\*1

P(F1=Yes|mammals) = 6 + 1 / 7 + 2

P(F2=No|mammals) = 6 + 1 / 7 + 2

P(F3=Sometimes|mammals) = 0 + 1 / 7 + 2

P(F4=Yes|mammals) = 5 + 1 / 7 + 2

P(F1=Yes|non-mammals) = 1 + 1 / 13 +2

P(F2=No|non-mammals) = 10 + 1 / 13 + 2

P(F3=Sometimes|non-mammals) = 4 + 1 / 13 + 2

P(F4=Yes|non-mammals) = 8 + 1 / 13 + 2

P1 = P(Mammals) \* P(F1=Yes|mammals) \* P(F2=No|mammals) \* P(F3=Sometimes|mammals) \* P(F4=Yes|mammals)

P1 = 8/22 \* 7/9 \* 7/9 \* 1/9 \* 6/9

= *.0163*

P2 = P(non-mammals) \* P(F1=Yes|non-mammals) \* P(F2=No|non-mammals) \* P(F3=Sometimes|non-mammals) \* P(F4=Yes|non-mammals)

P2 = 14/22 \* 2/15 \* 11/15 \* 5/15 \* 9/15

= *.0123*

% of P1 class = .0163 / .0163 + .0123 = 56.2%

% of P2 class = .0123 / .0163 + .0123 = 43.0%

**We choose P1**

**=> Mammals.**

See the attached Jupiter notebook for code

1. For the dataset -2 (liver disease dataset), implement Naïve Bayes classifier using Python. **[15M**]

Please follow the following steps.

1. Import the libraries [1M]
2. Load the dataset [1M]
3. Remove/replace missing values (if any) [2M]
4. Split features and labels [1M]
5. Split train and test data [1M}
6. Implement Naïve Bayes Classifier [2M]
7. Calculate accuracy measures [2M]

Submission Details

Text answers – id\_classification\_assignment2.doc

Code - id\_NB.ipynb

**Answer 3.** See the attached Jupiter notebook.