CSCI-5822 Probabilistic Models of Human and Machine Intelligence
Kumar Bhargav Srinivasan kusr7198@colorado.edu

Assignment 0

23rd January 2018

Tasks

1. Task 0
   Joint probability using 5 variable (two for class, one each for gender, age and outcome)
   Out of 32 possible entries, 24 are non zero and 8 combination of occurrences are not present in the events.

2. Task 1
   Build a probability table using Joint probability by using probability of
   $P(death|Gender = A, Age = B, Class = C)$ for all possible values of A,B,C
   For example to fill in
   $P(death|Gender = female, Age = adult, Class = 1st) =$
   $\frac{P(Gender = female, Age = adult, Class = 1st, Outcome = death)}{(P(Gender = female, Age = adult, Class = 1st, Outcome = death) + P(Gender = female, Age = adult, Class = 1st, Outcome = survival))}$

   ```
   probability table :
   Gender     female              male
   Age         adult     child     adult     child
   Class
   1st     0.027778  0.000000  0.674286  0.000000
   2nd     0.139785  0.000000  0.916667  0.000000
   3rd     0.539394  0.548387  0.837662  0.729167
   crew    0.130435       NaN  0.777262       NaN
   ```

   ```
   Classification Table :

   Gender       female                    male
   --------------------------------------------------
   Age      adult      child          adult   child
   --------------------------------------------------
   1st    Survival     Survival     Death    Survival
   2nd    Survival     Survival     Death    Survival
   3rd     Death        Death       Death    Death
   crew    Survival    undefined    Death    undefined
   ```

   Classification rule:
   From the probabilty table we could write a rule as :
   if conditional probability of given instance is nan then the prediction for the occurrence cannot be determined.
   if conditional probability of given instance is greater than 0.5 then the prediction would be death otherwise survival.
   this basically denotes that the classifier would classify to the class which provides maximum conditional probability : $argmax_c(P(Outcome|Gender, Age, Class))$, commonly called as Maximum

likelihood decision rule

Implementation:

Using pandas crosstab to find the frequency count irrespective of outcome and for specific outcome = death, following probability table was constructed using the frequency counts.

3. Task 2

   Build a naive bayes classifier for $P(death|Gender, Age, Class) = \frac{P(Gender|death) * P(Age|death) * P(Class|death) * P(death)}{(P(Gender|death)*P(Age|death)*P(Class|death)*P(death)+P(Gender|survival)*P(Age|survival)*P(Class|survival) * P(survival))}$
   using bayes theorem and conditional probability.

   Classification rule would be predict the class which is most probable, using the following equation : $argmax_c(P(Outcome) * P(Outcome|Gender, Age, Class))$, Also called as MAP decision rule - maximum a posteriori.

Gender	female		male	
Age	adult	child	adult	child
1st	0.0992700624909	0.0437273138056	0.527924239304	0.316934542432
2nd	0.206055613808	0.0972135395501	0.724782003544	0.522135194983
3rd	0.352318431243	0.184136083355	0.846617081228	0.696059298936
crew	0.367950929943	0.194547766253	0.855221720696	0.710219133754

   Following is the classification table :

Gender	female		male	
Age	adult	child	adult	child
1st	survival	survival	death	survival
2nd	survival	survival	death	Survival
3rd	survival	survival	death	death
crew	survival	survival	death	death

4. Task 3

   Pros and Cons of Empirical Joint probability vs Naive bayes Classification.

   (a) Empirical models can deduce any kind of relation between but without sufficient data they fail to predict output.
   (b) Joint probability models can easily overfit as all they predict directly based on the observed probabilities
   (c) Hard to compute Joint probability for higher n, as there will be 2^n possibilities.
   (d) Naive Bayes classifier assumes statistical independence of variables and hence allows to compute independent conditional probability and derive the prediction, which is an approximations which works most of the times, but not always.
   (e) By considering prior, we are able to make prediction for unseen combinations of the data.
   (f) Joint probability need to be regularized to reduce overfit whereas The prior beliefs about the parameters that they are driven by random data.
   (g) Interestingly with infinite data both them are able perform at same accuracy.