CIS – 490: Machine Learning
Learning Activity 1
Name: Pranav Vinod
UMassD ID: 01984464

Part 1

Q1. Describe the definition of Machine Learning.

Ans Machine learning can be described as a branch of Artificial Intelligence that using data and algorithms, based on statistical methods, automatically detects patterns in such data and makes predictions based on the discovered patterns.

Q2. Describe Statistical Learning and its overlap with Machine Learning.

Ans Murphy describes statistical learning as the vast set of tools to understand data. One might understand it as the statistical methods to draw information from data.
There are two kinds of statistical learning:

1. Supervised Statistical Learning – where we have both an input and output measurement
2. Unsupervised Statistical Learning – where we only have an input measurement and no output measurement.

Machine Learning could be understood as statistical learning done by a machine. With machine learning we try to make predictions or infer from a set of data to get information about some output variable.

Some common statistical distributions

1. Discrete Distributions:

a. Bernoulli Distribution

pmf = 1 – p if k = 0
p if k = 1

pmf = 1 – p if k = 0
p if k = 1

cdf = 0 if k < 0
1 – p if 0 ≤ k < 1
1 if k > 1

Application: In experiments and clinical trials, Bernoulli distribution is sometimes used to model a single individual experiencing an event like death, a disease or disease exposure.

b. Binomial Distribution

pmf = $C_k^n p^k q^{n-k}$;

where p is the success probability for each trial,
and q = 1 – p

cdf = $\sum_{i=0}^{|k|} \binom{n}{k} \; p^i q^{n-i}$;

|k| is the greatest integer less than or equal to k,
p is the success probability for each trial,
q = 1 – p

Application: Banks use the binomial distribution to model the probability that a certain number of credit card transactions are fraudulent.

c. Poisson Distribution

pmf = $\frac{\lambda^k \, e^{-\lambda}}{k!}$;

where λ is a parameter > 0
k is the number of occurrences

cdf = $e^{-\lambda} \sum_{i=0}^{|k|} \frac{\lambda^i}{i!}$;

where |k| is the greatest integer less than or equal to k,

Application: Call centers use Poisson distribution to model the number of expected calls per hour that they'll receive and that helps them to manage employee schedule based on number of calls they'll receive.

2. Continuous Distributions:

   a. Uniform Distribution

pdf = $^{1}/_{(b-a)}$ *for a ≤ x ≤ b*

0 *for x < a and x > b*

cdf = 0 *for x < a*

$^{(x-a)}/_{(b-a)}$ *for a ≤ x ≤ b*

1 *for x > b*

Application: The probability of a single number on the rolling of an unbiased die follows a uniform distribution because each number is equally likely to occur.

b. Gaussian / Normal Distribution:

pdf = $\frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}({x-\mu}/_{\sigma})^2}$;

where μ is the mean of the distribution,
and σ is the standard deviation of the distribution

cdf = $\frac{1}{\sqrt{2\pi}}\ \int_{-\infty}^{x} e^{-t^2/2}dt$

Application: A random sample of height of people follows the normal distribution.

c. Student t Distribution:

pdf = $\frac{1}{\sqrt{\nu}\mathrm{B}(\frac{1}{2},\frac{\nu}{2})}\ (1+\frac{t^2}{\nu})^{-(\frac{\nu-1}{2})}$;

where ν are the degrees of freedom,
and B is the Beta function

cdf = $1 - \frac{1}{2}\mathrm{I}_{x(t)}(\frac{1}{2},\frac{\nu}{2})$;

where I is the regularized incomplete beta function, and $x(t) = \frac{\nu}{t^2+\nu}$

Application: Student t distribution has important applications in hypothesis testing for small samples.

d. <u>Chi-squared Distribution:</u>

pdf = $\frac{1}{2^{k/2}\,\Gamma(\frac{k}{2})}\ x^{\frac{k}{2}-1} e^{-x/2}$;

where *k* is the degrees of freedom,
and $\Gamma(\frac{k}{2})$ is Gamma function

cdf = $\frac{\gamma(\frac{k}{2},\frac{x}{2})}{\Gamma(\frac{k}{2})}$;

where γ is the lower incomplete gamma function

Application: This distribution is often encountered in magnetic resonance imaging.

e. <u>Gamma Distribution:</u>

pdf = $\frac{x^{\alpha-1}\ \ e^{-\beta x}\ \ \beta^{\alpha}}{\Gamma(\alpha)}$;

where α is called a shape parameter,
and β is called a rate parameter,
and $\Gamma(\alpha)$ is the gamma function

cdf = $\frac{\gamma(\alpha,\beta x)}{\Gamma(\alpha)}$;

where γ is the lower incomplete gamma function

Application: The amount of rainfall in a reservoir are modelled after a gamma process.

f. Beta Function:

pdf = $\frac{x^{\alpha-1} \quad (1-x)^{(\beta-1)}}{\mathrm{B}(\alpha,\beta)}$;

where α, β are shape parameters > 0,
and $\mathrm{B}(\alpha, \beta)$ is the beta function

cdf = $\frac{\mathrm{B}(x{:},\alpha,\beta)}{\mathrm{B}(\alpha,\beta)}$;

where $\mathrm{B}(x{:}, \alpha, \beta)$ is the incomplete beta function

Application: The beta distribution can be used to model anything that has a limited range like from 0 to 1. It also has uses in order statistics.

g. Pareto Distribution:

pdf = $\frac{\alpha x_m^{\alpha}}{x^{\alpha-1}}$ $\quad x \geq x_m,$

0 $\quad x < x_m$

Where x_m is the minimum possible value of x, And α is a positive parameter

cdf = $1 - (\frac{x_m}{x})^{\alpha}$ $x \geq x_m,$

0 $x < x_m$

Application: This distribution is used to describe the allocation of wealth among individuals.

3. <u>Random Variable</u>: A random variable is a variable whose values are the result of a random experiment.

   There are two types of random variable:

   1. Discrete Random Variable – only takes discrete specific values
   2. Continuous Random Variable – can take any value in a continuous range.

<u>PART 2</u>

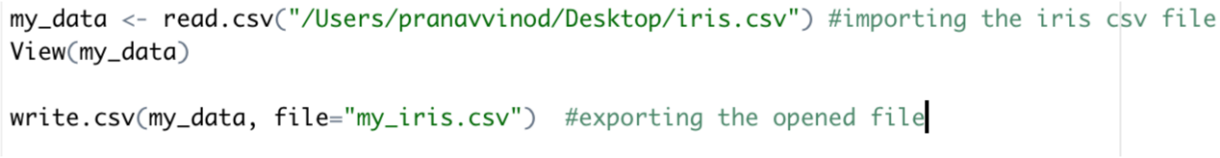
Q1. Import and Export iris dataset.

Ans

```
my_data <- read.csv("/Users/pranavvinod/Desktop/iris.csv")

View(my_data)

write.csv(my_data, file="my_iris.csv")
```



```
my_data <- read.csv("/Users/pranavvinod/Desktop/iris.csv") #importing the iris csv file
View(my_data)

write.csv(my_data, file="my_iris.csv")  #exporting the opened file
```

References:

1. https://www.statology.org/binomial-distribution-real-life-examples/
2. *https://www.statology.org/poisson-distribution-real-life-examples/*
3. https://www.statology.org/uniform-distribution-real-life-examples/
4. https://en.wikipedia.org/wiki/Chi-squared_distribution#Occurrence_and_applications
5. https://en.wikipedia.org/wiki/Gamma_distribution#Occurrence_and_applications