Assignment 1 (Sol.)

Introduction to Machine Learning
Prof. B. Ravindran

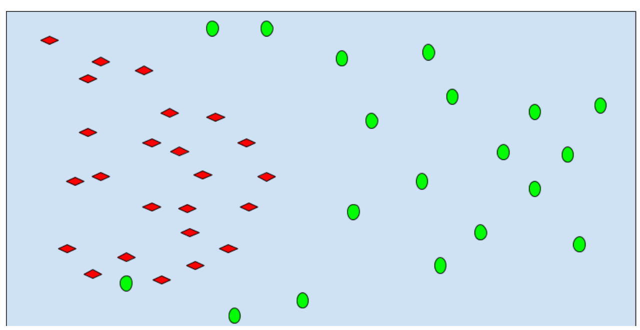
1. Which of the following tasks can be best solved using Clustering.

   (a) Predicting the amount of rainfall based on various cues
   (b) Detecting fraudulent credit card transactions
   (c) Training a robot to solve a maze

   Sol. (b)
   We can think of the task of detecting fraudulent credit card transactions as essentially representing all credit card transactions using some features and performing clustering. The majority of the transactions will be legal and come under one (or perhaps more) cluster whereas we hope to find transactions not in the above cluster to indicate fraudulent activity which can be detected.

   Predicting the amount of rainfall is essentially a supervised learning problem. On the other hand, training a robot to solve a maze would best be attempted by making use of reinforcement learning algorithms.

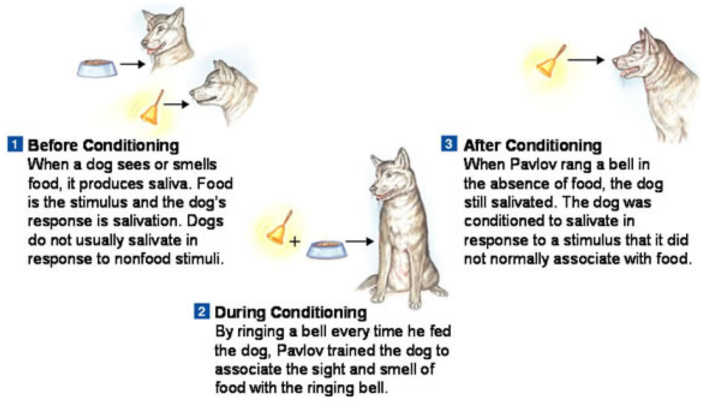
2. What would be the ideal complexity of the curve which can be used for separating the two classes shown in the image below.



   (a) Linear
   (b) Quadratic
   (c) Cubic

   Sol. (b)
   For the 2D data points shown in the figure, the distribution of the data suggests that the ideal complexity of the curve used for separating the two classes is quadratic. A linear boundary would result in a large number of miss-classifications, whereas a third degree curve would not do any better than the quadratic.

3. **Pavlov's Experiment**



(Image source - http://www.goldiesroom.org/)

Pavlov's experiment is a classic experiment conducted by the Russian physiologist Ivan Pavlov. He experiments on dogs shown in the image below.

Before conditioning dog responds with saliva only in presence of the food but after conditioning it starts salivating just with the bell. Select the correct option(s) about the experiment.

(a) In this experiment, the dog acts as a Reinforcement learning agent
(b) In this experiment, the dog learns in a supervised setting
(c) Comparing this experiment to Reinforcement learning theory, the various states are
  - Presence of just the bell
  - Presence of just food
  - Presence of both food and bell

Sol. (a) & (c)
In this scenario, we see that as the dog interacts with the environment it is in, it starts to modify its behaviour based on rewards it receives from the environment (in this case, food). This is clearly a reinforcement learning situation. We also note the different states of the environment which differ as the experiment progresses - initially just the food is present resulting in the dog producing saliva, next the bell is introduced whenever the dog is given food as part of the conditioning process, and finally, the desired behaviour (i..e, salivation) can be produced when just the bell is rung (without the food).

4. There are n bins of which the kth contains $k-1$ blue balls and $n-k$ red balls. You pick a bin at random and remove two balls at random without replacement. Find the probability that:
   - the second ball is red;
   - the second ball is red, given that the first is red.

(a) 1/3, 2/3
(b) 1/2, 1/3
(c) 1/2, 2/3
(d) 1/3, 1/3

Sol. (c)
Let C_i be the colour of the ith ball. In each bin, there are a total of $(k-1)+(n-k)=(n-1)$ balls. Of these half are blue and the other half are red (verify $\sum_{k=1}^{n} k-1 = \sum_{k=1}^{n} n-k$).

The probability of the second ball being red is equal to the probability of the second ball being red given that the first ball was either red or blue.

For a particular bin we have,

$$P(C_2 = red) = \frac{(n-k)}{(n-1)}\frac{(n-k-1)}{(n-2)} + \frac{(k-1)}{(n-1)}\frac{(n-k)}{(n-2)} = \frac{n-k}{n-1}$$

Considering all bins, we have

$$P(C_2 = red) = \sum_{k=1}^{n} \frac{n-k}{n(n-1)} = \frac{1}{2}$$

The probability of the second ball being red given that the first ball was red,

$$P(C_2 = red|C_1 = red) = \frac{P(C_2 = red, C_1 = red)}{P(C_1 = red)}$$

Now, $P(C_1 = red) = \frac{1}{2}$.
For a particular bin,

$$P(C_2 = red, C_1 = red) = \frac{(n-k)}{(n-1)}\frac{(n-k-1)}{(n-2)}$$

Considering all bins, we have

$$P(C_2 = red|C_1 = red) = \sum_{k=1}^{n} \frac{\frac{(n-k)(n-k-1)}{n(n-1)(n-2)}}{\frac{1}{2}}$$

Simplifying, we have

$$P(C_2 = red|C_1 = red) = \frac{2}{3}$$

5. A medical company touts its new test for a certain genetic disorder. The false negative rate is small: if you have the disorder, the probability that the test returns a positive result is 0.999. The false positive rate is also small: if you do not have the disorder, the probability that the test returns a positive result is only 0.005. Assume that 2% of the population has the disorder. If a person chosen uniformly from the population is tested and the result comes back positive, what is the probability that the person has the disorder?

   (a) 0.803
   (b) 0.976
   (c) 0.02
   (d) 0.204

Sol. (a)
Let,

- T be the probability of the test being positive,
- D be the probability of a person having the disorder

From the data provided, we have:

- $P(T|D) = 0.999$
- $P(T|\neg D) = 0.005$
- $P(D) = 0.02$
- $P(\neg D) = 1 - P(D) = 0.98$

We want to calculate the probability of a person chosen uniformly at random having the disorder given that the test came back positive, i.e.,

$$P(D|T)$$

Now from the Bayes' theorem, we have

$$P(D|T) = \frac{P(T|D)P(D)}{P(T)} = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|\neg D)P(\neg D)}$$

Substituting known values, we have

$$P(D|T) = \frac{0.999 * 0.02}{0.999 * 0.02 + 0.005 * 0.98} = 0.803$$

6. In an experiment, n coins are tossed, with each one showing up heads with probability p independently of the others. Each of the coins which shows up heads is then tossed again. What is the probability of observing 5 heads in the second round of tosses, if we toss 15 coins in the first round and $p = 0.4$?

   (Hint: First find the mass function of the number of heads observed in the second round.)

   (a) 0.372
   (b) 0.055
   (c) 0.0345
   (d) 0.0488

   Sol. (b)
   The same result will be observed if we toss each of the coins twice and count the number of coins showing two consecutive heads. The probability of showing two consecutive heads, given that the probability of showing heads in a toss is p, is equal to p^2. Hence, the number of heads X observed after the second round of tosses can be given by

$$P(X = r) = \binom{n}{x} p^{2r}(1 - p^2)^{n-r}$$

   Now substituting the given values into the above pmf we have,

$$P(X = 5) = \binom{15}{5} 0.4^{10}(1 - 0.4^2)^{10} = 0.055$$

7. Consider two random variables X and Y having joint density function $f(x, y) = 2e^{-x-y}, 0 < x < y < \infty$. Are X and Y independent? Find the covariance of X and Y.

   (a) Yes, 1/4
   (b) Yes, 1/2
   (c) No, 1/4
   (d) No, 1/2

   Sol. (c)
   From the joint density, we can compute the marginal densities of X and Y.

$$f_X(x) = \int_{y \geq x} 2e^{-x-y} dy = -2e^{-x}e^{-y}|_x^\infty = 2e^{-2x}.$$

$$f_Y(y) = \int_{x \leq y} 2e^{-x-y} dy = -2e^{-y}e^{-x}|_0^y = 2e^{-y}(1 - e^{-y}).$$

   We can see that $f_X(x)f_Y(y) \neq f(x, y)$, hence, X and Y are not independent.
   Now,

$$E[X] = \int_{x \geq 0} 2xe^{-2x} dx = \frac{1}{2}$$

$$E[Y] = \int_{y \geq 0} 2ye^{-y} dy - \int_{y \geq 0} 2ye^{-2y} dy = 2 - \frac{1}{2} = \frac{3}{2}$$

   Calculating the covariance, we have,

$$E[XY] = \int_{0 \leq x \leq y} 2xye^{-x-y} dxdy = \int_{x \geq 0} 2xe^{-x} dx \int_{y \geq x} ye^{-y} dy$$

   Now,

$$\int_{y \geq x} ye^{-y} dy = xe^{-x} - \int_{y \geq x} -e^{-y} dy = (x + 1)e^{-x}$$

   Thus, we have,

$$E[XY] = \int_{x \geq 0} 2x(x + 1)e^{-2x} dx = \int_{x \geq 0} 2x^2e^{-2x} dx + \int_{x \geq 0} 2xe^{-2x} dx$$

   Note that the second term is simply 1/2 whereas the first term

$$\int_{x \geq 0} 2x^2e^{-2x} dx = -\int_{x \geq 0} 4x(-(1/2)e^{-2x}) dx = \int_{x \geq 0} 2xe^{-2x} dx = \frac{1}{2}$$

   Combining all results, we have

$$Cov(X, Y) = E[XY] - E[X]E[Y] = 1 - \frac{3}{4} = \frac{1}{4}$$

8. An airline knows that 5 percent of the people making reservations on a certain flight will not show up. Consequently, their policy is to sell 52 tickets for a flight that can hold only 50 passengers. What is the probability that there will be a seat available for every passenger who shows up?

   (a) 0.5101
   (b) 0.81
   (c) 0.6308
   (d) 0.7405

   Sol. (d)
   The probability that there will be a seat available for every passenger who shows up is equal to the probability that less than or equal to 50 passengers show up. This is the same as 1 minus the probability that exactly 52 or 51 passengers show up. Thus, the required probability

$$= 1 - 0.95^{52} - 52(0.95)^{51}(0.05) = 0.7405.$$

9. Let X have mass function

$$f(x) = \begin{cases} \{x(x+1)\}^{-1} & \text{if x = 1, 2, ...,} \\ 0 & \text{otherwise,} \end{cases}$$

   and let $\alpha \in \mathbb{R}$. For what values of α is it the case that $\mathbb{E}(X^\alpha) < \infty$?

   (a) $\alpha < \frac{1}{2}$
   (b) $\alpha < 1$
   (c) $\alpha > 1$
   (d) $\alpha > \frac{3}{4}$

   Sol. (b)
   We have

$$E[X^\alpha] = \sum_{x=1}^{\infty} \frac{x^\alpha}{x(x+1)}$$

   This expression is finite only if $\alpha < 1$.

10. Is the following a distribution function?

$$F(x) = \begin{cases} e^{-1/x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

   If so, give the corresponding density function. If not, mention why it is not a distribution function.

   (a) No, not a monotonic function
   (b) Yes, $x^{-2}e^{-1/x}, x > 0$
   (c) No, not right continuous
   (d) Yes, $x^{-1}e^{-1/x}, x > 0$

Sol. (b)
$F(x) \to 1$ as $x \to \infty$ and $F(x) = 0$ for $x < 0$ by definition. The monotonicity and continuity of F follow from the corresponding properties of $e^{-1/x}$. Also, since all continuous functions are right continuous, F is right continuous. Thus, the given function is a distribution function. Its corresponding density function is given by differentiating:

$$f(x) = \begin{cases} x^{-2}e^{-1/x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$