

# COMP4702/COMP7703/DATA7703 - Machine Learning

## Homework 0 - Background Knowledge Review

Marcus Gallagher

Due: 12:00pm Tuesday 5th March, 2019

### Task

Machine Learning is a subject which draws on specific background knowledge from a variety of different disciplines, including computer science, statistics and mathematics. The purpose of this task is to provide some examples of useful background knowledge for machine learning. If you have the recommended background knowledge, this may be a refresher for you. It is still possible to successfully complete the course if you do not already have all the background knowledge, as long as you are prepared to review this knowledge as it becomes useful during the course.

### Questions

1. Find the sample mean and (unbiased) sample variance of the following data set: (to 2 decimal places)

$$X = [13, 15, 24, 32, 4, 9, 10, 14, 18, 19]$$

2. If  $M$  is a  $3 \times 5$  matrix,  $R$  is a  $5 \times 7$  matrix and  $S$  is a  $7 \times 2$  matrix, how many rows and columns are in  $(M \times (R \times S))$ ?
3. Given the following three 5-D vectors:

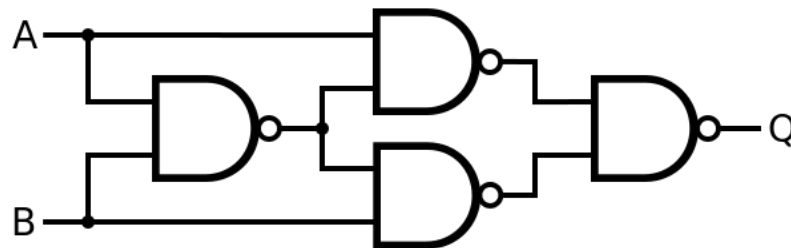
$$\mathbf{w} = [0.4201, 0.8402, 0.2801, 0.1400, 0.1400], \mathbf{u} = [0.1348, 0.4045, 0.5394, 0.2697, 0.6742], \\ \mathbf{v} = [0.5330, 0.3198, 0.2132, 0.7462, 0.1066]$$

- (a) Calculate the dot product of the two most orthogonal vectors. (to 4 decimal places)
  - (b) Calculate the angle between these two vectors. (in degrees and to two decimal places)
4. Given the following matrix, answer the questions below.

$$\mathbf{M} = \begin{bmatrix} 9 & 2 & 4 & 1 \\ 3 & 5 & 6 & 9 \\ 2 & 1 & 7 & 8 \\ 6 & 3 & 7 & 4 \end{bmatrix}$$

- (a) What is the largest eigenvalue of  $M$ ? (round to 2 decimal places)

- (b) What is the determinant of  $M$ ?
- (c) What is the product of the eigenvalues? (round to the nearest whole number)
5. The `rand()` function in Matlab produces a pseudo-random number that is (choose the correct answer):
- (a) Between -1 and 1
- (b) Between 0 and 1
- (c) Drawn from a Gaussian distribution with mean = 0 and variance = 1
- (d) An integer
6. Determine the output ( $Q$ ) of the following logic circuit, given the following input: ( $A = 1, B = 0$ )



Note: The only gate included in the circuit above is the NAND gate, which behaves as follows:

A	B	Q
0	0	1
1	0	1
0	1	1
1	1	0

7. Select the correct set of inputs and outputs for the following (Matlab) function:

---

```

function output = example(arr , n)
output = [];
idx = n;
while length(output) ~= length(arr)
    while idx > length(arr)
        idx = idx - length(arr);
    end
    disp(idx);
    output = cat(2,output , arr(idx));
    n = n + 1;
    idx = idx + n;
end

```

---

- (a)  $arr = [1, 2, 4, 0, 7]$ ,  $n = 3$ ,  $output = [4, 2, 2, 4, 7]$
- $arr = [6, 3, 5, 1]$ ,  $n = 4$ ,  $output = [1, 6, 5, 3]$
- $arr = [4, 3, 8, 6]$ ,  $n = 5$ ,  $output = [4, 6, 3, 8]$

(b) arr = [3, 2, 4, 6, 7], n = 3, output = [4, 2, 2, 4, 7]  
 arr = [6, 3, 5, 1], n = 8, output = [1, 6, 5, 3]  
 arr = [5, 1, 6, 8], n = 5, output = [5, 6, 1, 1]

(c) arr = [4, 2, 4, 2, 7], n = 3, output = [4, 2, 2, 4, 7]  
 arr = [6, 3, 5, 1], n = 4, output = [1, 6, 5, 3]  
 arr = [8, 2, 6, 5], n = 5, output = [8, 6, 2, 6]

8. Given the following data:

x	y	z
3	4	1
6	7	3
1	4	8
9	3	2
1	6	5
2	2	2
1	7	6
4	5	4

What is the Pearson correlation coefficient between the two columns with the highest correlation (in absolute value)? (to 2 decimal places)

9. Correctly order ((a) through (h)) the following computational complexities from (asymptotically) slowest to fastest:

- (a)  $O(n^3)$
- (b)  $O(1)$
- (c)  $O(\log(n) + n)$
- (d)  $O(n \log(n))$
- (e)  $O(\log(n))$
- (f)  $O(n^2 + n \log(n))$
- (g)  $O(n!)$
- (h)  $O(2^n)$