

# **Statistical Machine Learning 2022**

**(COMP4670 / 8600)**

## **Take-Home Practice / Hurdle Exam *Solutions***

*Writing period : 90 minutes.*

*Study period : 0.*

*Permitted materials : A computer for the coding question only.*

*Total marks: 17.*

***This does not contribute to your grade. However if you find it at all difficult to score fifty percent without any preparation then you should seriously consider dropping the course.***

1. (1 mark)

In a box there are two black balls, one red ball and three yellow balls. What is the probability that a ball selected at random is yellow given that it is not black?

**Solution:**

(2 half marks)  $3/4$ .

2. (1 mark)

Which of the following statements is correct?

i)  $p(a,b)/p(b) = p(b|a)$

ii)  $p(a|b)/p(a) = p(b|a)$

iii)  $p(a,b)/p(b) = p(a|b)$

iv) None of the above.

**Solution:**

(2 half marks)  $p(a,b)/p(b) = p(a|b)$

3. (1 mark)

A fair six-sided die has 1, 1, 2, 3, 5 and 8 dots on each face. What is the expected number of dots?

**Solution:**

(2 half marks)  $(1 + 1 + 2 + 3 + 5 + 8)/6 = 3\frac{1}{3}$ .

4. (2 marks)

Let  $D$  be a random variable with smooth (infinitely differentiable) probability density function on the real line. Let  $x$  and  $y$  be independent and identically distributed according to  $D$ . What is the probability that  $x > y$ ?

i) 0

ii)  $1/2$

iii) 1

iv) It depends on the choice of  $D$ .

v) None of the above.

**Solution:**

(4 half marks)  $1/2$ .

5. (1 mark)

What is the derivative of  $x^x$  with respect to  $x$ ?

**Solution:**

(2 half marks)  $x^x(1 + \log(x))$

6. (2 marks)

Let  $f(x) = \frac{1}{2}ax^2 + bx + c$ , where  $a < 0$ . What is the maximum value obtained by this function, for real valued  $x$ ?

**Solution:**

(4 half marks)  $-\frac{1}{2}b^2/a + c$ .

7. (1 mark)

What is the order of computational time complexity (roughly speaking, the number of computational operations required) of matrix-vector multiplication, assuming the matrix is of size  $N$  by  $N$ ?

- i)  $O(1)$
- ii)  $O(N)$
- iii)  $O(N^2)$
- iv)  $O(N^3)$
- v)  $O(N \log N)$
- vi) None of the above.

**Solution:**

(2 half marks)  $O(N^2)$ .

8. (2 marks)

Consider a linear regression model  $y = ax + b$  with  $(x, y)$  data pairs  $(1, 2), (1, 3), (2, 3)$ . What is the derivative with respect to  $a$  of the mean squared error, at  $a = 1, b = 2$ ?

**Solution:**

(4 half marks)  $1/2$ .

9. (1 mark)

If  $A$  is a square matrix, then we write  $\text{tr}(A)$  for its trace, which is the sum of the diagonal elements of  $A$ . Let  $\|A\|_F = \sqrt{\text{tr}(A^\top A)}$  be the Frobenius norm of a matrix.

Let  $C$  be a fixed symmetric  $n \times n$  matrix (so  $C = C^\top$ ). Let  $\mu$  be a scalar that is larger than the  $p^{\text{th}}$  smallest eigenvalue of  $C$ . Let  $N$  be a diagonal  $p \times p$  matrix with distinct positive entries on the diagonal.

Let our cost function  $f(X)$  be defined for  $n \times p$  matrices  $X$  as

$$f(X) = \frac{1}{2} \text{tr}(X^\top C X N) + \mu \frac{1}{4} \|N - X^\top X\|_F^2,$$

where  $X \in \mathbb{R}^{n \times p}, n \geq p$ .

Derive the gradient  $\nabla_X f(X)$ .

**Solution:**

(2 half marks) See the labs.

10. (1 mark)

What is the name of the numpy function that samples from the “standard normal” distribution?

**Solution:**

(2 half marks) `numpy.random.randn`

11. (4 marks)

Complete the implementation of  
`linear.py` released with this problem set,  
in order to successfully run the program

test\_linear.py.

Which you may execute using *e.g.* the command

anaconda3-python test\_linear.py

**Solution:**

(8 half marks) `w = np.linalg.pinv(np.dot(x, x.T)) @ np.dot(x, y)`