

Student Name: Saeedreza Zouashkiani

Student ID: 400206262

Deep Learning Homework 1

# 1)

## 1-a)

## 1-b)

### 1-b-i)

Since X is non-negative then is positive, and is larger than . Thus:

Therefore

### 1-b-ii)

### 1-b-iii)

Let denote the random variable (indicator) that determines whether the random point falls within the circle. can take a value of with probability of and with a probability of . Therefore

The estimator then is . We check whether is an unbiased estimator of .

### 1-b-iv)

Using Chebyshev’s inequality, we have:

Then if we solve for , we get . Therefore

# 2)

## 2-i)

## 2-ii)

By rearranging the terms and multiplying by from left we get

## 2-iii)

Let be the (i, j) minor of, (i, j) element of cofactor matrix of , and adjugate matrix of respectively.

By cofactor expansion of **A**

# 3)

Let be an matrix.

## 3-i)

The characteristic polynomial of A is defined as

Also, the characteristic polynomial can be factorized as

So, by comparing terms we get

## 3-ii)

By comparing terms from last part, it is easily derived.

# 4)

## 4-i)

If A has full column rank, the inverse of exists

To check whether is a pseudoinverse, we should have:

Or if we have

## 4-ii)

If **A** has full row rank, the inverse of exists

Or

# 5)

## 5-i)

We start by eliminating the block matrix under A

Then by eliminating the element above D

**=**

Therefore, by combining the two

Thus

Which is the LDU decomposition of M.

Therefore

## 5-ii)

Like the last part, using block LDU decomposition when D is invertible, we get

## 5-iii)

Assume that **A** and **D** are both invertible, therefore by inverting the LDU decomposition of part i and ii we get. First invert using part i.

Now inverting part ii, results in:

By comparing the first block matrix we get

If we substitute with

## 5-iv)

Using part I, with :

# 6)

## 6-i)

Using question 5

## 6-ii)

Using Courant-Fischer’s theorem

To give an upper bound on a minimum value of a function, we just need to give an upper bound on some value it takes.

Let be subspaces of with dimensions of and respectively which achieve the minimum values of , and let be their intersection. has dimension of at least .

Since has dimension of at least , the above is an upper bound on the value of for any dimensional subspace

# 7)

Differentiating with respect to theta and setting to zero, yields:

# 8)

## 8-i) ML

Taking derivative of log likelihood function with respect to and setting it to zero

## 8-ii) MAP

We want to maximize

Taking the log

Taking derivative with respect to and setting it to zero

# 9)

Let so that

We will at first prove part ii

## ii)

Let be a subset matrix of dimension such that , if the element in corresponds to element in , and zero otherwise.

Therefore, by applying linear transformation

Therefore is a normal distribution with

## i)

The joint distribution of and is and from part ii, the marginal distribution of ­is . According to Bayes’ law

Denote the inverse as

According to question 5

Plugging the inverse into (1):

Also using the determinant derived in question 5

Thus

Which is the probability distribution of multivariate normal distribution

# 10)

## 10-i)

The last term does not affect the minimization over x. By differentiating with respect to x and setting to zero:

Therefore

## 10-ii)

Assume that the algorithm has converged. We shall have

## 10-iii)

Let be the optimum solution. By adding and subtracting from both sides (Using t’th iteration notation by subscript)

Define

We know that can be diagonalized by . Multiply both sides by and define .

is a diagonal matrix. Therefore, for the i’th element of after k iterations

Therefore, for the i’th mode to converge ( we shall have

And to satisfy convergence of all modes we shall have