IML Exercise 2 Answers

1. Theoretical part
   1. Solutions of The Normal Equations
2. We want to show that   
   We’ll use SVD of

So by the SVD of we can see that

Since the vector has 0’s in the first positions, and has 0’s in the columns onwards:

On the other hand:

Looking on the same we saw that when we calculated , so now:

Note SVD is a single decomposition so it’s the same that we got for .

1. We want to prove that for any square matrix :

We’ll show first:

i.e. we’ll show:

We’ll show the complement:

By showing:

means it has a component of i.e.

Since and we know and it’s also true specifically for

Norm is 0 only for so

1. For a non-homogeneous system where is a square matrix but noninvertible, we want to show that we have solutions

We know the square matrix is noninvertible only if it has linear dependency in it’s rows. If has a matching dependency in it’s elements like we have in than we have solutions. Otherwise, the non-homogeneous equation has 0 solutions.

When has a matching dependency in it’s elements like we have in : so

We saw in previous question (2) that it implies so

The way back is also true since all steps we used were “”

1. We’ll show that for the linear system (where is the variable), we have either a single solution or solutions.

If is invertible, we have a solution: and we have a single solution.

If is not invertible, if we have a trivial solution, otherwise:

Note Q2 conclusion: is as

By Q3 fitted to our eq’: has solutions

So we conclude we have solutions

* 1. Projection Matrices

1. For (outer product),
   1. P is symmetric:

Observe the Matrices

And similarly, in

* 1. e.val of P are 0 or 1, and e.vac of e.val 1 are :

For

So we saw are e.vecs with e.val 1

The base can be orthogonally expanded to by when each vectors set so:

For

So we saw are e.vecs with e.val 0

* 1. :

Every can be expressed by the vectors that spans the sub-space

When because the diagonal has only either values or 0s or 1s

* 1. :

* 1. Least Squares

1. We’ll show  when is invertible

We’ll use SVD of   
Note that is of full rank with non-zero singular values, otherwise would not have a full rank and was not invertible. And we proved

Define

Assumed invertible:

When

1. We’ll show is invertible :

We saw in Q1 that and note

is invertible means

Rank of in means it spans all .

All the way back it true as well so its ‘’

1. We’ll show that when is **not** invertible, :

Like in Q1 we’ll look on the SVD of , it has the same kernel as meaning they are both not invertible and their has singular values of zero

We’ll use SVD of

Separating the first rows/columns which match non-zero singular values we can write it also in sub-matrices format:

The last equivalence is because is orthogonal matrix of rotations only and doesn’t change the norm.

This is minimized when: i.e.

Since

we concluded

This has values of 0’s in indexes because other values in these indexes will give valid solution but will enlarge the that we wanted to minimize.

1. Practical part
   1. Univariate Gaussian Estimation
2. From a 1000 samples of normal distribution of with np.random.seed(0), we got estimated mean , variance (unbiased estimator) of:

(9.954743292509804, 0.9752096659781323)

Calculated by:

1. When sample set size is increasing from 10 to 1000 only on the samples set we already took on Q1, the consistency is demonstrated.

Text

Description automatically generated with low confidence

1. The Probability-Density-Function of the values in the data set of 1000 samples of with np.random.seed(0), is compared here versus the ideal PDF.

We can see we under-estimated the variance (0.975) and the ideal was slightly higher (1.0)

The PDF of the sample points are on the estimated normal distribution model.

Chart, line chart

Description automatically generated

* 1. Multivariate Gaussian Estimation

1. From a 1000 samples of normal distribution of when

with np.random.seed(0),

We estimate expected value by:

We estimate variance value by:

we got estimated mean , variance (unbiased estimator) of:

Estimated mu vector is

[-0.02282878 -0.04313959 3.9932571 -0.02038981]

Estimated Covariance (Sigma) matrix is

[[ 0.91667608 0.16634444 -0.03027563 0.46288271]

[ 0.16634444 1.9741828 -0.00587789 0.04557631]

[-0.03027563 -0.00587789 0.97960271 -0.02036686]

[ 0.46288271 0.04557631 -0.02036686 0.9725373 ]]

1. With the same covariance Matrix and same samples as in Q4, we scan the most probable vector.

We expect to get the result of and this is indeed the point with the highest log-likelihood.

1. The highest probable are

(f1,f3) = (-0.05, 3.97)  
(Note it comes from an estimation with resolution of and at the center of the range we have ~ … -0.15, -0.05, 0.05, 0.15 … )