Exercise 2

Machine Learning I

|  |  |
| --- | --- |
|  | 2A-1. |

1. Ordinary Multiplication is not defined for vectors. The dot product is no ordinary multiplication, as it does not satisfy e.g. field axioms.

|  |  |
| --- | --- |
|  | 2A-2. |

## a)

We want to decompose an arbitrary matrix into:

where is a symmetric matrix and is skew symmetric.

We basically have the constraints:

Setting

satisfies the system of equations and maintains the symmetric properties. This decomposition is comparable to the Euler decomposition of the complex sine and cosine functions.

## b)

Just plug the previously found solution into the polynomial:

## c)

In a symmetric matrix, each row has independent entries.

Consequently, the number of entries is equivalent to the sum of natural numbers up until But this sum can already be expressed in closed form by the well-known Gauss formula:

|  |  |
| --- | --- |
|  | 2A-3. |

|  |
| --- |
| Auxiliary calculation |

|  |
| --- |
| Inverse of a 2x2 Matrix:  Additionally, explicit calculation for with gives: |

Let be two dimensional. Recovery of by marginalizing out:

Let . Then:

Let us isolate the first exponent and simplify it:

Continuation of previous marginalization:

The general proof looks like that:

|  |
| --- |
| Prerequisites |

|  |
| --- |
| Given with the following block structure:  where , .  As seen in [1], every matrix that is in the same equivalence class as the matrix below, is a valid inverse if are Hermitian (i.e. symmetric if matrices are real):  Note: denotes the Hermitian transpose. Also:  Additionally, we decompose data into:  If you compare sums, you will notice:  For any vector with aforementioned dimensions. |

Let us first focus on **:**

Now that the exponents are separated, they can be factored out of the integral:

This can now be normalized by the appropriate factors so that the integral is and our resulting distribution is normal.

|  |  |  |
| --- | --- | --- |
|  |  | 2A-4. |

Let the be defined as:

We try to find values for which the gradient vanishes. Afterwards, we verify if the found solutions are local/global maxima.

|  |
| --- |
|  |

Because is strictly positive on , the position of extreme values is invariant under logarithmic transformations. This helps to facilitate easier differentiation, as products turn into sums:

Calculation of

Calculation of

To show that these values indeed maximize is left as an exercise to the reader (just take the Hessian and see if ).

For given data , we can visually inspect the validity of the alleged extrema.

|  |  |
| --- | --- |
|  |  |
| Visualization of for the dataset  and , .  The sample and variance averages  , indeed represent the maxima. | On the unconstrained domain reaches two symmetric maxima. The restriction removes negative values of . |

Conveniently the estimators that maximize are the sample variance and sample mean. This sample variance is biased. The unbiased estimate would be , see Bessel’s Correction.

Please remember, existence of partial derivatives does not imply that is differentiable. But since and are also continuous, is differentiable.

|  |  |
| --- | --- |
|  | 2A-5. |

### Lazy version:

If we assume that is already normal, we only need to find the mean vector and covariance matrix:

### Complete version:

|  |
| --- |
| Prerequisites |

|  |
| --- |
| We require the following properties of matrices :  All three properties are satisfied by any invertible complex matrix.  Additionally, we use the change of variable formula:  If  then: |

In this case we have:

This can now be plugged into the pdf of