Abstract

This document contains relevant information about the IML exam.

Exam Preparation

Dan Witzner Hansen

December 2023

Figure 1: Map of the IML content.

Exam

The exam is individual, oral (zoom) and will last up to 30 minutes.

1 Before Exam

We urge you to thoroughly prepare for the exam. This includes

1. Read all material, tutorials and watch the video material (the curriculum). **Please be aware that the curriculum may include course material not encompassed by the exercises.**

2. Prepare and practice a presentation for each exam question, complete with key points to focus on (as listed below). You are permitted to bring a summary outline and your Jupyter notebooks/examples to the exam, which

you can use to illustrate outcomes and discuss their significance. While you may display the outcomes of your exercises, refrain from demonstrating or executing code during the exam unless specifically requested—your time is better allocated to discussing the issue at hand, its theoretical context, formal models, the data, and the findings. Feel free to incorporate both "important" and "extra" exercises into your exam presentation.

2 At the Exam

Enter the room Be ready in in front of the the examination room (or in zoom waiting room) at least 30-60 minutes before your scheduled time.

Draw Question (0-1') You will randomly select a question corresponding to one of the weekly exercises or mandatory assignments. Each question has multiple focus points, and one will be chosen at random for you.

Present Question (1-6') In 5-6 minutes, present the exercise, including an overview, purpose, main findings, and its theoretical context. Focus on the most substantial part of the exercise with the given focus points, demonstrating depth and relevance to the course material.

Answer General Questions (6-25') Initially, we will not interrupt your presentation, but we may ask clarifying questions. As time progresses, we will pose questions about your presentation, theory, and other related topics within the curriculum.

Evaluation (25-30') After about 20-35 minutes, your direct involvement ends, and you will be asked to wait outside the room (in the waiting room if online). We will then invite you back to inform you of your grade. Detailed feedback may be scheduled for another day upon request.

3 Change of Schedule

While it is uncommon, please be aware that exam dates could potentially be adjusted up to one week before the scheduled exam day. It is also essential that you check the most up-to-date exam schedule on the morning before your exam to confirm any recent changes.

If a student is absent on the day of the exam, we will proceed by calling the next student in line. Please be aware that there may be delays, and as a result, the scheduled exam times might shift throughout the day. Therefore, we ask that you be available and ready during the morning or afternoon session (at least 30-60 minutes before), as per your assigned schedule. **To benefit all students, it is crucial to maintain an accurate exam schedule. Therefore, if you foresee any circumstances, such as illness, that might prevent you from attending your exam, please notify Dan as early as possible. Prompt communication will enable timely adjustments to the**

schedule, ensuring minimal disruption and maintaining fairness for all students.

Questions

The bullet points listed for each week denote the specific focus areas for the corresponding questions. At the exam, one of these bullet points will be selected at random to determine your precise focus for discussion.

Q1 Exercises Weeks 2 (Vectors)

(a) Focus on inner products, vector operations, distance metrics and their relation to ML (evaluation and other metrics). You may relate this to week 9 but focus on vectors.

Q2 Exercises Week 3 (Matrices)

(a) Focus on matrices and their operations (addition, multiplication, transpose. inverse, determinant, orthogonal). Relate them to their application within the course (e.g., transformations, basis).

(b) Focus on how linear equations and their solutions are related to matrices.

Q3 Exercise Week 4 (Linear Transformations)

(a) Focus on linear transformations in 2D and 3D but relate to linear transformations of higher dimensions and non-linear transformations.

(b) Focus on affine transformations, homogeneous coordinates and composition of linear transformations.

Q4 Exercise Week 5 (Least Squares)

(a) Focus on the relation between linear least squares (function minimization) and projections

(b) Focus solving linear least squares problems for model fitting (including design matrix)

(c) focus on Mandatory 1

Q5 Exercise Week 6 (Undestanding data, Data Cleaning, Uncertainty)

(a) Uncertainty / descriptive statistics. Relate this to model learning and data evaluation.

(b) Data, uncertainty and over/underfitting.

(c) Focus on uncertainty, data cleaning in relation to regression, classification, clustering or dimensionality reduction.

(d) Missing data, duplicate data, outlier detection, and data imputation.

Q6 Exercise Week 7 (Regularization, Filtering)

(a) Focus on filtering (convolution and correlation, noise, image gradients)

(b) Focus on matching and metrics.

(c) Focus on Cross validation

(d) Focus on Regularization

(e) Focus on Bias /Variance, R^2

Q7 Exercise Week 8 (Classification)

(a) Linear classification and "kernels"

(b) Logistic regression

(c) Features/HOG Features and classification

Q8 Exercise Week 9 (Evaluation)

(a) Metrics/Evaluation of Classifiers

(b) Metrics/Evaluation of Regression models

(c) Features/HOG Features and classification.

(d) Imbalanced data for classification and regression

Q9 Exercise Week 10 (Principal Component Analysis)

(a) Dimensionality reduction and PCA. Focus on mandatory 2

(b) Generative models and PCA

(c) Eigenvalues, covariance matrix and basis

Q10 Exercise Week 11 (Clustering and Non-linear optimization)

(a) K-means and Mean shift

(b) K-means and Algomerative clustering

(c) Kmeans and ELBOW

(d) Nonlinear functions, graphs, gradients and gradient descent with relation to model training and non-linear models

Q11 Exercise Week 12 (Neural networks)

(a) Neural networks prediction (regression vs classification)

(b) Neural networks training (Gradients, the chain rule and back/forward propagation)

(c) Training and Evaluation

Q12 Exercise Week 13 / 14 (Architectures)

(a) Model architectures: Difference between fully connected /multi layer perceptron (MLP) and CNN

(b) Regularization, data augmentation, model complexity and norms

(c) Model tuning, dropout, early stopping, complexity

(d) Convolutional Neural Networks (CNN) (Convolutional layers, max pooling)

Additional preparation

The subsequent subsections present reflective questions designed to engage you with key elements of the course. This document is evolving and may be updated. While not exhaustive, these questions offer a means to contemplate various aspects of the course content. They are phrased broadly to facilitate a wide range of reflections.

4 Linear Algebra

Basics

- List as many purposes for which we use vectors for image analysis and Machine learning
- What is the equation of a line, planes and hyperplane using vector notation?
- How do you calculate the length and orientation of a vector ?
- How do we know when two vectors u and v are orthogonal to each other.
- How do we know when two vectors u and v are parallel to each other?

Linear Equation

What is a linear equation and how is this relate to matrices ?

Inner product

How is the inner product related to:

- a measure of distance.
- matrix multiplication
- projections
- convolution
- neural networks

Solutions to Linear Equation

- What does it mean to have a solution to a linear set of equations?
- When can we have one, zero or many solutions to a linear set of equations?

- What is an over-determined set of equations.
- What is an under-determined set of equations.
- Why is the Determinant relevant when talking about solutions to linear equations.
- Why are subspaces important when talking about solutions to linear set of equations.
- Given data $X \in \mathbb{R}^N$ and labels y. How w can linear equations be used to find the coefficients of the following models and how much training data is needed to learn the model parameters
  (a) A straight line in the plane
  (b) A plane in 3D
  (c) A hyperplane in N-dimensional spaces
  (d) Find the coefficients of a an N-order polynomial
  (e) Find the coefficients of an similarity or affine transformations
- In the above cases what is the minimal number of points needed to solve the linear set of equations.

Transformation

- What is a transformation and how is it related to a projection.
- Matrix multiplications may be considered as a transformation. Why?
- How are linear transformations combined?
- What is the purpose of homogeneous coordinates.
- What is the inverse of a transformation and what is its relevance to the course / ML.
- How is least squares (formally) related to projections

5 Signals

- What are the definitions of convolution and correlation and how are they related
- When can correlation and convolution be used interchangeably
- How can correlation be implemented in a neural architechture and why is this beneficial?
- How is image templates useful as a machine learning model and a metric for comparison.
- How do image templates relate to machine learning

6 Machine Learning

- Where is supervised and unsupervised learning used and how do they differ.
- How is least squares used in machine learning and how does it relate to least squares when using matrices.
- What is an objective function
- Which methods can be used to learn linear and non-linear models
- What is regularization and why is it needed.
- How do recommender systems work and how is this related to inner products and matrix factorization.
- What is PCA? and how does PCA make use of subspaces , eigenvalues and eigenvectors.
- How is PCA and certain neural architectures related.
- Why is it called linear classification
- How does logistic regression differ from linear classifiation.
- What is the decision boundary and how can you find it.
- what is a kernel, where are they used and how is it related to model learning

7 Evaluation

- Why is evaluation needed
- How do you ensure proper evaluation of models
- Why are training, test and verification sets needed in the training procedures?
- What is cross validation and how is it related to overfitting / underfitting
- How how can we tell when a model is under and overfitted?