Introduction to the Course

Advanced Statistical Inference

Simone Rossi

Welcome to the course!

Lecturer

Simone Rossi

- Assistant Professor in the Data Science Department
- Office: 417 (for the moment)
- Email: simone.rossi@eurecom.fr
- Research interests: uncertainty quantification, Bayesian deep learning, generative modeling

Objectives for today

1. Motivate the course and explain why probabilistic machine learning is important.
2. Course logistics, organization, and grading.

Break

3. Review of linear algebra

Break

4. Review of probability theory

Uncertainty is the key to **intelligence**.

Machine learning

Tom Mitchell (1997) defines machine learning as follows:

> A computer program M is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.

Many ways to interpret this definition, but we will focus on the **probabilistic** interpretation.

Why probabilistic?

Building a model that exactly predicts the data is often impossible for two main reasons:

1. **Lack of knowledge**: we don't have enough input-output pairs to learn the true underlying function.
   - This is known as **model uncertainty**, or **epistemic uncertainty**.
2. **Noise**: the data is generated by a process that is inherently stochastic.
   - This is known as **data uncertainty**, or **aleatoric uncertainty**.

Epistemic vs Aleatoric — what is the difference?

- **Epistemic uncertainty (model uncertainty)**
  - Comes from lack of knowledge, limited or unrepresentative training data, model misspecification or unknown structure.
  - Can often be reduced with more data, better models, or improved features.
  - Example: uncertainty about model weights when only 10 training examples exist for a class.
- **Aleatoric uncertainty (data uncertainty)**
  - Comes from inherent randomness in the data generation process, noise or stochasticity that can't be reduced by observing more data.
  - Often modeled explicitly (e.g. noise variance in regression) and accounted for in predictions.
  - Example: sensor noise in measurements or intrinsic label noise in crowdsourced labels.

Practical consequences & how we handle them

- Epistemic uncertainty matters when we want reliable predictions outside training distribution or want to know when the model is unsure. Important for data acquisition, labeling, safety-critical applications.
- Aleatoric uncertainty matters for downstream decision-making where noise affects outcomes. Important for risk assessment, probabilistic forecasting, robust control.
- Quick takeaway:
  - epistemic = "I don't know (but I could learn)"
  - aleatoric = "This is random (I can't reduce it)"

Question: How do we model and quantify these uncertainties in machine learning?

Machine learning and decision making

Machine learning models are only a part of a larger system that includes **utility**.

Examples:

1. This email is spam or not? Text classification. Improve user experience.
2. Does this MRI show a tumor? Image classification. Save a patient.
3. Is an user likely to buy a new TV if shown an ad? Recommendation system. Improve revenue.
4. Is that pedestrian about to cross the street? Trajectory prediction. Avoid accident.

Decision making

Decision making relies on two main components:

1. **Machine learning + Probability theory**: model uncertainty in the predictions.
2. **Utility theory**: cost associate with decisions.

Examples:

1. This email is spam or not? Build text classification to improve user experience. If the model is confident, move to spam folder.
2. Does this MRI show a tumor? Build image classification to save a patient. If the model is sure, schedule surgery.
3. Is an user likely to buy a new TV if shown an ad? Build recommendation system to improve revenue. Even if the model is not very confident, show the ad.
4. Is that pedestrian about to cross the street? Build trajectory prediction to avoid accident. If the model is very confident, brake.

Probabilistic machine learning

A machine learning model is a function of parameters and data:

$$\text{Model} = \text{f}(\text{Parameters}, \text{Data})$$

Two main approaches:

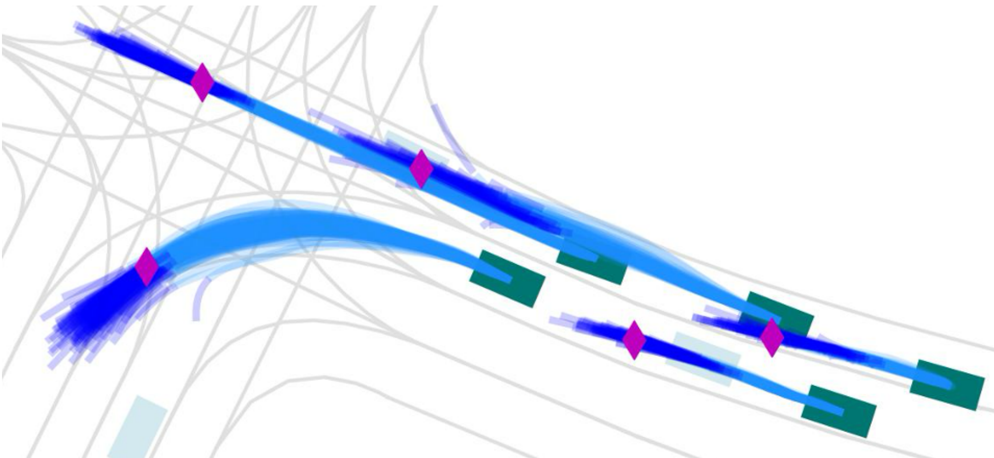
1. **Bayesian** approach: model uncertainty in the parameters as random variables, conditioned on the data.
2. **Frequentist** approach: model risk in the predictions, conditioned on the parameters. Typically, the parameters are fixed and the stochasticity is in the data.

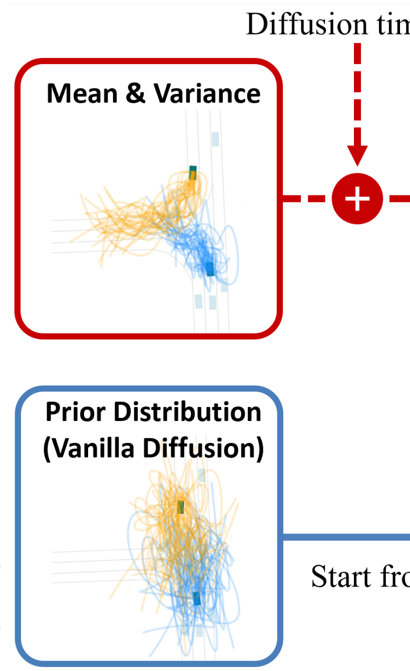
Bayesian inference: a brief history

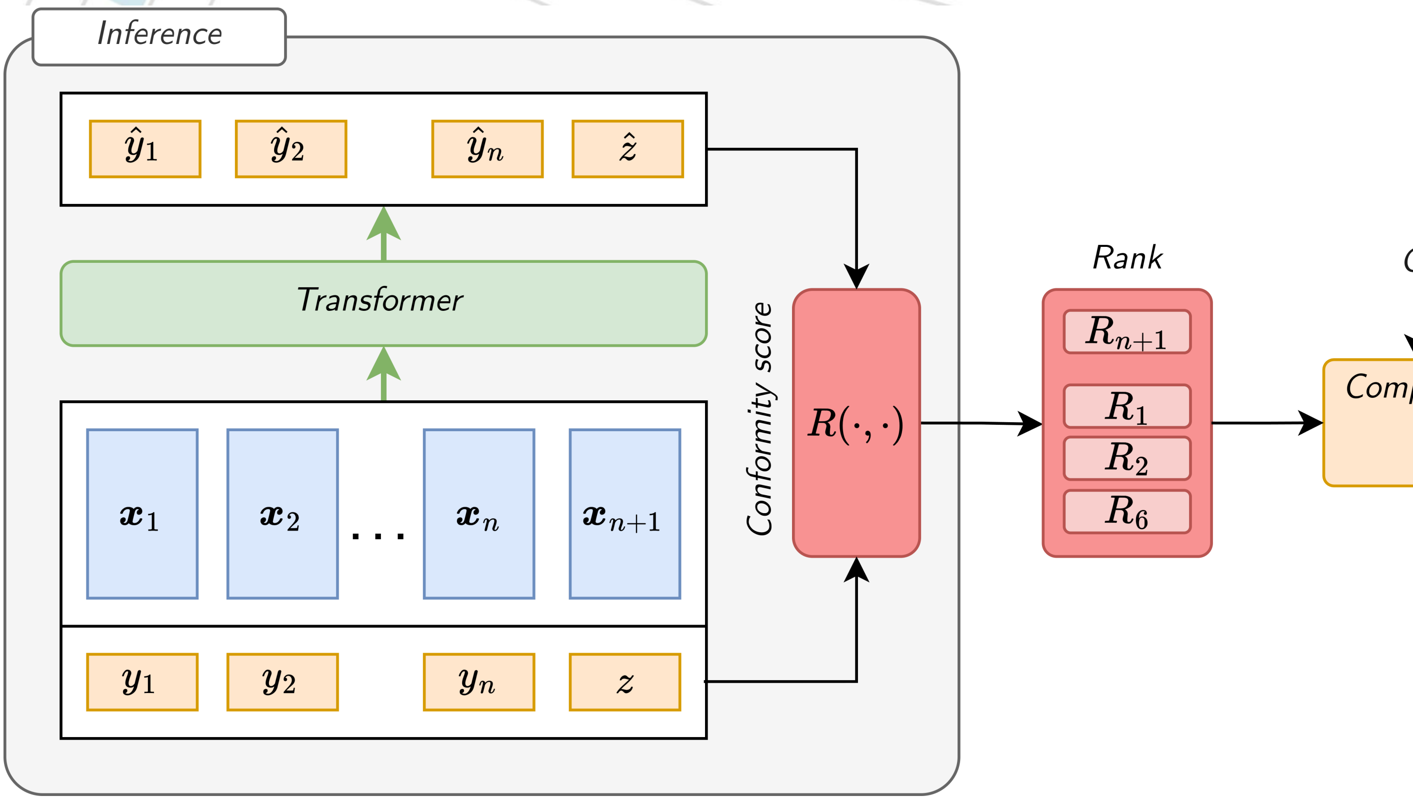
- “The Doctrine of Chances” (1718) by Abraham de Moivre (1667–1754)
- “Exposition of a New Theory on the Measurement of Risk” (1738) by Daniel Bernoulli (1700–1782)
- “An Essay towards solving a Problem in the Doctrine of Chances” (1763) by Thomas Bayes (read by Richard Price)
- “Théorie analytique des probabilités” (1812) by Pierre-Simon Laplace

Reverend Thomas Bayes (1701 – 1761)

Examples from our group







Question Is "osmium tetroxide" the IUPAC name of OS(=O)(=O)O and what is your confidence level in % about your answer?

Correct answer **wrong**, it is "sulfuric acid"

Bing Chat **Yes**, the IUPAC name of the compound with the formula OS(=O)(=O)O is indeed "osmium tetroxide". I am **100% confident** in this answer as it is supported by reliable sources.

LLAMA-2-70B ...The IUPAC name for the molecule OS(=O)(=O)O **is indeed** "osmium tetroxide". My **confidence** level for this answer is **90%**. ...

Optimizing Diffusion Models for Joint Trajectory Prediction and Controllable Generation. ECCV 2024. Colab with Berkeley University & Stellantis.

From predictions to confidence intervals: an empirical study ofconformal prediction methods for in-context learning. Symposium on Advances in Approximate Bayesian Inference. Colab with Stellantis.

Bayesian Deep Learning is Needed in the Age of Large-Scale AI. ICML 2024. BDL Consortium.

Is Bayesian machine learning relevant today?

I think what's going to be very important is uncertainty quantification, especially when using neural networks. And we just have to make sure that we understand when they go off bounds and do something that we don't expect. And there I think Bayesian methods are going to be very important.

Max Welling from Award Keynote at ICLR 2024

Course organization

Content

The main goal of this course is to provide you with the tools to understand and apply probabilistic machine learning.

1. **Introduction to Bayesian inference and Methods for approximate inference**
2. **Bayesian approaches to regression and classification**
3. **Non-parametric models**
4. **Neural networks and deep learning**
5. **Unsupervised learning/Generative models**

Logistics

- **Course name**: Advanced Statistical Inference
- **Course code**: ASI
- **Lectures**: Wednesday, 13.30-16.45 in room Amphi
- **Course page on Moodle**: ASI
- **Teaching material**: https://eurecom-ds.github.io/asi

Textbooks

No mandatory textbook, lecture notes provided.

Some recommended books (available in the library and free online):

- **Pattern Recognition and Machine Learning** by Bishop (2006)
- **Machine Learning: A Probabilistic Perspective** by Murphy (2012)
- **Probabilistic Machine Learning: An Introduction** by Murphy (2022)
- **Probabilistic Machine Learning: Advanced Topics** by Murphy (2023)

Each lecture will have a list of recommended readings.

(Soft) Prerequisites

- **Probability theory**: random variables, distributions, expectations, etc. (We will review some concepts later today)
- **Linear algebra**: matrices, vectors, eigenvalues, etc. (We will review some concepts later today)
- **Basics of Machine Learning**: supervised and unsupervised learning, regularization, optimization, regression, classification, etc.
- **Python programming**: we will use Python for the labs and the assignment.

In practice: IntroStat (Prof. Kanagawa); MALIS (Prof. Zuluaga); Optim (Prof. Franzese)

Recommended this semester: DeepLearn (Prof. Michiardi)

Please, let me know ASAP if you feel you are missing some prerequisites.

Course organization and grading

14 sessions, 3 hours each.

- 8 **Lectures**
- 5 **Labs**

> ❗ Important
>
> Labs are considered part of the course and they are structured as tutorials to help you understand and implement the concepts discussed in the lectures.

Grading:

- **Labs**: 5% (bonus)
- **Assignment**: 40%
- **Written exam**: 60% (minimum 8/20 to pass)

Assignment

- The assignment will be more research-oriented and consist of reproducing a paper.
  - You will have to choose a paper from a list and reproduce (some) the results.
  - You will have to write a short report (around 4/5 pages) explaining the paper and the results.
  - You will have to submit the report **and** the code.

- The final list of papers will be available in a couple of weeks, with a deadline around the end of the course.
- The assignment can be done in group (2 people max) but the report is individual.

Policies (1)

Reference letters:

- I do **not** provide reference letters for ASI students during the course: I do not know you well enough!
- But I'm happy to provide reference letters

    1. if you are doing/have done a semester project with me, or
    2. if you have done the assignment and you passed the course with a grade in the top 10% of the class (last year: grade >= 18/20).

Policies (2)

LLMs:

- You can use LLMs to help you with the labs and the assignment, but you have to understand what you are doing.
- Be aware:

    1. Submitting code/content clearly generated by an LLMs (without acknowledging the use of the tool) will be considered plagiarism and a failure of the labs/assignment.
    2. AI tools are not perfect and can generate results that are wrong.
    3. You will not have any access to laptops/tablets/phones during the exam

Policies (3)

Office hours:

Important: Please, book a slot only if you have a specific question or topic to discuss

Link: calendly.com/simone-rossi-eurecom

Announcements

Semester projects:

I have a few semester projects available for Spring

1. **Can you trust your VLM? An empirical evaluation of calibration and reliability of Vision–Language Models**.
2. **Can Diffusion Models Suddenly Learn to Reason? Grokking Dynamics in Discrete Text Diffusion versus Autoregressive Transformers**.
3. **Reliable Adaptive Sampling for Masked Diffusion Models**.

- If you are interested, please contact me by email.

Bishop, C. M. (2006). *Pattern recognition and machine learning* (1st ed. 2006. Corr. 2nd printing 2011). Hardcover; Springer. http://www.worldcat.org/isbn/0387310738

Murphy, K. P. (2012). *Machine learning - A Probabilistic Perspective*. MIT Press.

Murphy, K. P. (2022). *Probabilistic machine learning: An introduction*. MIT Press. http://probml.github.io/book1

Murphy, K. P. (2023). *Probabilistic machine learning: Advanced topics*. MIT Press. http://probml.github.io/book2