# Topics 2022

Topics for the course **CS-E4875 Research Project in Machine Learning, Data Science and Artificial Intelligence** for the academic year 2022-2023.

Examples of the previous year: Topics 2021– read only

**USE THIS FIRST TOPIC AS A TEMPLATE AND COPY-PASTE-EDIT IT TO YOUR NEEDS:**

**Topic #0: TITLE**

**Background:**

**Prerequisites:**

**Language requirements:**

**Supervisor name and email:**

**Supervisor's research group leader's name unless supervisor him/herself is one:**

**Topic available:** no **|** yes_one_instance |  yes_many_instances (max M instances)

**Topic available also for a group:** no | yes (max N students in the group)

---

**Topic #1: Physics-based Character Animation with Sampling-based Trajectory Optimization**

**Background:** Physics-based character animation is a viable testbed for artificial intelligence techniques with many exciting downstream applications such as robotics, films, games, and many more. There are two main branches of character control via virtual motors equipped to the character's joints: closed-loop (e.g. reinforcement learning) and open-loop (e.g. trajectory optimization), both of which can leverage physics simulation to a large extent. We investigate the latter, implementing some of the state-of-the-art black-box optimization techniques in the form of gradient descent and/or expectation maximization among other possibilities. We will design and deploy interesting characters and scenarios and visualize diverse and emergent behaviours, analyzing the characters' capabilities to perform the tasks at hand. As a reach goal, we will push the envelope of trajectory optimization by introducing novel techniques emerging from our investigation.

Take a look at these works for some visual examples:

http://www.cs.utoronto.ca/~mazen/spacetime/

http://libliu.com/Samcon2/Samcon2.html

https://dl.acm.org/doi/10.1145/3197517.3201315

https://dl.acm.org/doi/10.1145/3072959.3073707

https://dl.acm.org/doi/10.1145/2767002

https://www.cs.ubc.ca/~van/papers/2021-SCA-modassist/index.html

**Prerequisites:** In this project, we optimize workflow for growth and effectiveness, not efficiency. Priority will be given to the candidate(s) demonstrating the strongest ability to work together, dive deep into research questions, and exercise ownership. The ideal candidate(s) will work closely with the supervisor, actively participating in the weekly literature review, writing, and pair-programming sessions. Candidates will be invited to an informal 1-hour pair-programming and learning session audition to try out the workflow before deciding to come on board. These backgrounds would be a bonus (and not necessary): rudimentary Python coding, some knowledge of reinforcement learning and optimization, some experience in graphics (e.g. Jaakko Lehtinen's undergraduate graphics course), some experience in working with visualization libraries (e.g. matplotlib).

**Language requirements:** English

**Supervisor name and email:**  Nam Hee Gordon Kim (namhee.kim@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:** Perttu Hämäläinen

**Topic available:** yes_many_instances (max 2 instances)

**Topic available also for a group:** yes (max 2 students in the group)

---

**Topic #2: Misinformation classification and community detection on Twitter**

**Background:** Acting on the wrong information can kill. According to the World Health Organization, nearly 6 000 people around the globe were hospitalized in the first 3 months of 2020 because of coronavirus misinformation. During this period, at least 800 people may have died due to misinformation related to COVID-19. At its extreme, death can be the tragic outcome of misinformation. False information runs the gamut, from discrediting the threat of COVID-19 to conspiracy theories that vaccines could alter human DNA and so on.

An ongoing project “Crisis Narratives”, multidisciplinary research consortium between Aalto University and Finnish Institute for Health and Welfare (THL) investigates Covid-19 related narratives. Part of this project, the task focuses on what kind of misinformation types shared during Corona epidemic, who shared this type of information, how online activism was organized, how this threat emerged and changed on Twitter.

**Prerequisites:** Basics of Data Mining, programming skills, preferably R/Python.

**Language:** Finnish (more preferable since the data is in Finnish language) or English

**Supervisor name and email:** Ali Unlu (ali.unlu@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:** Nitin Sawhney

**Topic available:** yes_one_instance

**Topic available also for a group:** yes (max 2 students in the group)

---

**Topic #3:** Approximate Bayesian inference in Bayesian Deep Learning

**Background:** The task of this project is to implement and evaluate recently developed inference methods from the statistics community to perform approximate inference in Bayesian deep learning models (e.g., Bayesian neural networks or Bayesian probabilistic circuits). The topic will be selected together with the student. Examples are: sampling with PDMPs in Bayesian neural networks, Bayesian bootstrap for Bayesian neural networks and comparison to deep ensembles, and non-reversible Langevin method for Bayesian probabilistic circuits.

**Prerequisites:** knowledge of Bayesian methods and programming experience in Julia or interest in learning Julia

**Language:** English

**Supervisor name and email:** Martin Trapp (martin.trapp@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:** Arno Solin

**Topic available:** yes_one_instance

**Topic available also for a group:** no

---

**Topic #4: Solving differential equations with equivariant neural networks**

**Background:** Physics-informed neural networks (PINN) [1] is a machine-learning way of solving partial differential equations: a neural network is trained to approximate the solution by minimizing the fit to the initial conditions and the loss derived from the form of the solved equations. In this project, we will train PINNs with a recently proposed architecture which has an in-built property of translation equivariance [2].
[1] M. Raissi, P. Perdikaris, G. E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational Physics 378 (2019) 686–707.
[2] T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, T. Aila (2021). Alias-free generative adversarial networks. Advances in Neural Information Processing Systems, 34, 852-863.

**Prerequisites:** Basics of deep learning (CS-E4890 Deep Learning), programming skills (Python),

**Language requirements:** English

**Supervisor name and email:** Alexander Ilin (alexander.ilin@aalto.fi) and/or Katsiaryna Haitsiukevich (katsiaryna.haitsiukevich@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:** Alexander Ilin

**Topic available:** yes_one_instance

**Topic available also for a group:** yes (max 2 students in the group)

---

**Topic: #5: Predicting patients’ length of stay in ICU**

**Background**: Electronic health records (EHR) contain information about individuals' visits to healthcare services. The student will get here a chance to wor k with MIMIC-IV data, which is a large, freely available database comprising deidentified health-related data. The goal is to predict patients' length of stay in ICU based on data, which is collected from their 24 first hours spent in ICU. We have cleaned a data set from thousands of individuals and implemented our first machine learning models. In this project, the student will implement an additional model to the task!

Link to the MIMIC-data: https://physionet.org/content/mimiciv/2.0/

An article that introduces benchmark data sets: https://www.nature.com/articles/s41597-019-0103-9.pdf

**Prerequisites**: Good programming skills (Python) and a solid background in machine learning

**Language**: Finnish or English

**Supervisor name and email:** Tommi Gröhn (tommi.i.grohn@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:** Pekka Marttinen

**Topic available:** yes_one_instance

**Topic available also for a group:** yes (max 2 students in the group)

---

**Topic #6: Sentiment Analysis and Federated Deep Learning**

**Background**: I am proposing two subtopics this year.

Subtopic 1: Deep learning for multimodal sentiment analysis, which aims to devise a novel multimodal fusion method for sentiment analysis with textual, visual, and acoustic data. This subtopic focuses on deep model fusion for multimodality learning and natural language processing techniques for textual modality.

Reference: https://arxiv.org/abs/2208.11893

Subtopic 2: A survey on learning algorithms in federated learning, which plans to review federated learning algorithms such as federated transfer learning, knowledge distillation, and unsupervised learning.

Reference: https://arxiv.org/abs/2102.12920

If you are looking for other similar tasks, check out the details about my research projects via this link https://users.aalto.fi/~jis1/supervision/

**Prerequisite**: 1) knowledge of deep learning; 2) programming skills with deep learning frameworks (e.g., PyTorch); 3) experience with Latex typesetting and Linux servers.

**Language requirements: English**

**Supervisor**: Shaoxiong Ji (shaoxiong.ji@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:** Pekka Marttinen

**Topic available:** yes_one_instance

**Topic available also for a group:** no

---

**Topic #7: Kinematic Character Animation with Intelligent Motion In-Painting**

**Background:** Learning to leverage rich motion capture datasets is a crucial part of modern intelligent animation systems. In this project, we will investigate data-driven keyframing as a special case of in-painting in the space-time domain. Consequentially, we apply modern neural networks-based techniques. We will produce rich visualizations showing what makes motion in-painting unique compared to other problems. We will analyze resulting animations and think about limitations and potential areas of improvement. As a reach goal, we will work towards pushing the envelope of data-driven keyframing and kinematic character animation with novel methods incorporating our insights.

https://www.cs.ubc.ca/~van/papers/2018-MIG-autocomplete/index.html

https://montreal.ubisoft.com/en/automatic-in-betweening-for-faster-animation-authoring/

https://ojs.aaai.org/index.php/AAAI/article/view/20368

https://ait.ethz.ch/projects/2020/motion_infilling/

**Prerequisites:** In this project, we optimize workflow for growth and effectiveness, not efficiency. Priority will be given to the candidate(s) demonstrating the strongest ability to work together, dive deep into research questions, and exercise ownership. The ideal candidate(s) will work closely with the supervisor, actively participating in the weekly literature review, writing, and pair-programming sessions. Candidates will be invited to an informal 1-hour pair-programming and learning session audition to try out the workflow before deciding to come on board. These backgrounds would be a bonus (and not necessary): rudimentary Python coding, some knowledge of deep learning and optimization, some experience in graphics (e.g. Jaakko Lehtinen's undergraduate graphics course), some experience in working with visualization libraries (e.g. matplotlib).

**Language requirements:** English

**Supervisor name and email:** Nam Hee Gordon Kim (namhee.kim@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:** Perttu Hämäläinen

**Topic available:** yes_many_instances (max 2 instances)

**Topic available also for a group:** yes (max 2 students in the group)

---

**Topic #8: Applying federated learning on speaker verification for privacy**

**Background:** In any speech interaction system one of the first tasks is to verify if a speech command can be accepted or needs to be rejected. This needs to be done as fast as possible, needs to be private and needs to be executed continuously.

It is important to make sure that private information can not be leaked during this process, as typically, more than three seconds of recorded audio is required to verify a speaker reliably.

This project implements a federated learning algorithm (one project) to retrain a verification network on the go, with unsupervised representation learning (one project).

The goal is to increase the privacy in shared speech content as well as decreasing the necessary recording time. This project can also be subdivided into its individual parts with a narrower focus.

References: https://research.aalto.fi/en/publications/the-use-of-audio-fingerprints-for-authentication-of-speakers-on-s

**Prerequisites:** Programming skills in Python, as well as a machine learning framework. Which, Pytorch or Tensorflow does not matter.

**Language requirements:** English

**Supervisor name and email:** Silas Rech (silas.rech@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:** Tom Bäckström

**Topic available:** yes_one_instances

**Topic available also for a group:** yes (max 2 students in the group)

---

**Topic #9: Active speaker detection using GNN**

**Background:** Graph Neural Networks (GNN) have been widely explored in many computer vision applications, to name a few, object detection, image and video understanding, text classification, and speech recognition. Recently, GNN-based model has been proposed for Active Speaker Detection (ASD) through learning Spatial-Temporal graphs to encode the inter-modal relationship between audio and visual data. The proposed approach modeled ASD problem as a node classification task, where nodes representation is a global model that recursively aggregates information from neighboring nodes. In fact, nodes are associated with different local contexts and usually inconsistently distributed over the graph. The goal of this project is to handle ASD problem with graph classification instead of node classification using GNN. It is interesting to explore graph classification for ASD and the robustness of GNN in learning spatial-temporal features. We expected that ASD accuracy will be improved as well as elevating complex structure in node representation learning.

Referencs:

https://github.com/SRA2/SPELL

https://github.com/fuankarion/active-speakers-context

Examples:

https://github.com/pyg-team/pytorch_geometric

**Prerequisites:** Good python programming. Proficiency in GNN learning is a plus.

**Language requirements:** English

**Supervisor name and email:** Jalil Saif (abduljalil.saif@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:** Jorma Laaksonen

**Topic available:** yes_one_instance

**Topic available also for a group:** no

---

**Topic #10: Learnable upsampling schemes for real time speech super resolution**

**Background:** Oftentimes audio containing speech is transmitted at sample rates such as 16kHz or lower. Even when it is considered that such rates contain all the necessary information to successfully establish communication between two or more subjects, sometimes it is desirable to artificially extend the bandwidth to improve the resulting perceptual quality of the speech. Such task is known as bandwidth extension or audio super resolution. This procedure involves an upsampling stage to increase the output size to the target sample rate. This is commonly achieved by using transposed convolutional layers. Due to the so called checkerboard artifacts generated by this type of layers, several learnable alternatives have been proposed that may improve the quality and computational efficiency of the result. The goal of this project is to investigate these alternatives within an existing bandwidth extension neural network to determine which one is more performant and efficient in a real time context.

References:

- Real-time speech frequency bandwidth extension
- Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network
- Upsampling artifacts in neural audio synthesis

**Prerequisites:**

- Basic knowledge of digital audio
- Familiarity with python and deep learning frameworks (preferably PyTorch)

**Language requirements:** English

**Supervisor name and email:** Esteban Gómez <esteban.gomezmellado@aalto.fi>, Mohammad Vali <mohammad.vali@aalto.fi>

**Supervisor's research group leader's name unless supervisor him/herself is one:** Tom Bäckström

**Topic available:** yes_one_instance

**Topic available also for a group:** yes (max 2 students in the group)

---

**Topic #11: Deep learning in speech and language processing**

**Background:** Deep learning has changed the ways how speech and language data can be processed and represented. Several specific topics are available either for experimenting with new deep learning architectures in real-word data or applications, such as automatic speech recognition, understanding, translation and language learning and modeling. The topic can be selected together with the student.

**Prerequisites:** Aalto's basic course in either speech recognition or natural language processing or corresponding knowledge. Knowledge in deep learning. Experience in scientific programming, e.g. in Python**.**

**Language requirements:** English

**Supervisor name and email:** Mikko Kurimo mikko.kurimo@aalto.fi, Tamas Grosz tamas.grosz@aalto.fi and the automatic speech recognition (ASR) research group

**Topic available:** yes_multiple_instances

**Topic available also for a group:** yes (max 2 in a group)

---

**Topic #12: Speech-based biomarking of the human health state with machine learning**

**Background:** Speech signal provides an attractive means to predict the human health state. Increasing research interest is devoted particularly to detect neurodegenerative diseases, such as Parkinson's disease and Alzheimer's disease, from speech signals using both classical ML methods (such as SVMs) and more recent deep learning methods. Specific topics are provided in this health -related research area at the Department of Signal Processing and Acoustics.

**Prerequisite:** Knowledge in machine learning and deep learning. Basic knowledge in speech processing is highly recommended. Experience in scientific programming**.**

**Language requirements:** English

**Supervisor name and email:** Paavo Alku (paavo.alku@aalto.fi), Kiran Reddy (kiran.r.mittapalle@aalto.fi), Sudarsana Kadiri (sudarsana.kadiri@aalto.fi)

**Topic available:** yes_multiple_instances

**Topic available also for a group:** yes (max 2)

---

**Topic# 13 : Transformer-based Extreme Classification for Text Data**

**Background:** Extreme Classification (XC) aims to model problems with possibly millions of outputs as a supervised text classification task. Some real-world applications of XC include recommending Amazon products, predicting tags for Wikipedia articles, and next-word prediction in language models. Rec ently, transformer-based NLP models, such as BERT, have been applied to XC leading to more accurate models. However, the high computational complexity due to the large label space is still a bottleneck to the scalability of these models. The project involves investigating different ways, such as different negative sampling methods, for reducing the cost of computing the loss and updating the parameters of transformer-based XC models during training.

**Related Work:**

- https://arxiv.org/pdf/2101.03305.pdf
- https://arxiv.org/abs/2110.00685

**Prerequisites:** Knowledge of deep learning, experience with deep learning frameworks (preferably PyTorch)

**Supervisor name and email:** Rohit Babbar (rohit.babbar@aalto.fi)

**Topic available:** yes_many_instances (max 2 instances)

**Topic available also for a group:** yes (max 2 students in the group)

---

**Topic #14: Applied Machine Learning**

**Background:** Machine learning (ML) methods learn a hypothesis out of a hypothesis space (or model) that minimizes the loss incurred in predicting the label of a data point from its features. This project allows you to practice the art of modelling some particular application (based on your hobbies, studies or research) as a ML problem. You will also train good practices for applying ML methods to your problem.

**Prerequisites:** High-school level mathematics (the concept of vectors, matrices and functions) and programming (preferably in Python) skills.

**Language requirements:** English

**Related Work:**

- A. Jung, "Machine Learning: The Basics," Springer, Singapore, 2022. https://primo.aalto.fi/permalink/358AALTO_INST/ha1cg5/alma999531020706526

**Supervisor name and email:** Alexander Jung (alex.jung@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:**

**Topic available:** yes_many_instances (max 100 instances)

**Topic available also for a group:** no

---

**Topic #15: Active speaker localization**

**Background:** Active speaker localization has become one of the promising methods for speech and video scene understanding. Deep learning-based techniques can be exploited for localizing the active speaker in video frames. For example, audio-guided visual attention

mechanism is a potential for active speaker localization.

**Prerequisites:** Good python programming. Proficiency in PyTorch is a plus.

**Language requirements:** English

**Supervisor name and email:** Jalil Saif (abduljalil.saif@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:** Jorma Laaksonen

**Topic available:** yes_one_instance

**Topic available also for a group:** no

---

**Topic #16:** Effect of turn condition in HMC-NUTS MCMC algorithm

**Background:** Dynamic Hamiltonian Monte Carlo with no-U-turn-criterion is the state-of-the-art Markov chain Monte Carlo method for Bayesian inference in widely used probabilistic programming frameworks such as Stan, PyMC, TensorFlow probabilities, and Pyro. HMC alterantes Hamiltonian dynamic simulation and refreshment of energy level. The optimality of no-U-turn-criterion has been derived conditionally on fixed energy level and with focus on central statistics. The project involves review of the method and experiments to quantify how much the overall efficiency can be improved by using a modified stopping criterion for the Hamiltonian dynamic simulation.

**Prerequisites:** Bayesian inference and MCMC. Preferably R skills. Use of Python is possible but more challenging.

**Language requirements:** English

**Supervisor name and email:** Aki Vehtari (aki.vehtari@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:**

**Topic available:** yes_one_instance

**Topic available also for a group:** yes

---

**Topic #17:** Hamiltonian Monte Carlo diagnostic via variation in target curvature

**Background:** Hamiltonian Monte Carlo variants are the state-of-the-art Markov chain Monte Carlo methods for Bayesian inference in widely used probabilistic programming frameworks such as Stan, PyMC, TensorFlow probabilities, and Pyro. The algorithm includes discretized simulation of Hamiltonian dynamics. An efficient discretized simulation is challenging if the curvature of the posterior distribution varies a lot. In the worst case, very high curvature compared to the discretization step size leads to diverging simulations and bias. The project involves a short review of how function and gradient evaluations can be used to estimate curvature, and experiments for how local curvature estimates can be used to diagnose the variability of the curvature and indicate problematic regions.

**Prerequisites:** Bayesian inference and MCMC. Preferably R skills. Use of Python is possible but more challenging.

**Language requirements:** English

**Supervisor name and email:** Aki Vehtari (aki.vehtari@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:**

**Topic available:** yes_one_instance

**Topic available also for a group:** yes

---

**Topic #18:** Conformal Bayesian computation with Pareto smoothed importance sampling

**Background:** Conformal inference can provide better calibrated predictive intervals in case of model misspecification. In case of Bayesian models and Markov chain Monte Carlo methods, add-one-observation-in importance sampling can be used for efficient conformal inference. The project involves short review of the method, implementation of the method in Stan probabilistic programming ecosystem, and making experiments and a case study to illustrate the usefulness of the approach for model checking.

**Prerequisites:** Bayesian inference and Monte Carlo. R skills.

**Language requirements:** English

**Supervisor name and email:** Aki Vehtari (aki.vehtari@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:**

**Topic available:** yes_one_instance

**Topic available also for a group:** yes

---

**Topic #19: Diffusion planning for compositional generalization**

**Background**: Recently there has been a lot of interest in diffusion models, especially because they seem to have some capability to create novel combinations of familiar concepts ("compositional generalization"). A diffusion approach for planning and control problems[1] has recently been published.

We want to apply this to a simple grounded language learning problem[2] and see if works on compositional plan generation problems. For example, if an agent knows how to "push a cylinder while spinning" and "pull a cylinder", can it also know how to "pull a cylinder while spinning". This is a challenging problem that strong baselines like Transformers cannot solve very well[3]. If using diffusion for these sorts of problems works well, there may be an opportunity to publish the results as part of a conference paper.

**Prerequisites:** You are comfortable working with python, pytorch and numpy. You have taken the deep learning course (CS-E4890) or have equivalent background. You have some familiarity with how diffusion and transformers work. You don't need to know anything about compositional generalization or grounded language learning, we will provide the necessary background.

**References**:

[1] Janner. M et al, 2022, Planning with Diffusion for Flexible Behaviour Synthesis https://arxiv.org/abs/2205.09991
[2] Ruis. L et al, 2020, A benchmark for systematic generalization in grounded language understanding https://arxiv.org/abs/2003.05161
[3] Qiu. L et al, 2021, Systematic generalization on gSCAN, what is nearly solved and what is next? https://arxiv.org/abs/2109.12243

**Language requirements:** English

**Supervisor name and email:** Alexander Ilin (alexander.ilin@aalto.fi) and Sam Spilsbury (sam.spilsbury@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:** Alexander Ilin

**Topic available:** yes_one_instance

**Topic available also for a group:** yes (max 2 students in the group)

---

**Topic #20: Smart decoding from transformers with tree-search algorithms**

**Background:** Typically language models like transformers decode "autoregressively", which means that the outputs are generated one token at a time, conditioned on all the previously generated tokens. This is a sensible approach, but its hard to control the kinds of outputs that will be sampled and optimize them at runtime to meet certain conditions. Recently there have been some proposed approaches for tree-search based decoding, where possible sequences to decode can be seen as sort of nodes in a tree, and decoding means searching the tree for the best node meeting some conditions [1][2].

We think this method of decoding might have some very useful applications, for example, improving the quality of machine translation or abstractive summarization by backtranslating and aligning outputs to see if they still match what was in the input sequence. In particular, we think that this can probably help a lot with "hallucination", which is a really important problem in the field of natural language generation[3].

**Prerequisites:** You are comfortable working with python, pytorch and numpy. You have taken the deep learning course (CS-E4890) or have equivalent background. You have some familiarity with how transformers work. It would also be helpful to have taken a course in NLP or NLG, such as the Statistical Natural Language Processing course (ELEC-E5550) or the Seminar in Natural Language Generation.

**References:**

[1] Lu. X et al, 2022, NeuraLogic A*-esque-decoding https://arxiv.org/abs/2112.08726
[2] Chaffin. A et al 2022, PPLM-MCTS https://arxiv.org/abs/2109.13582
[3] Ziwei J. et al 2022, Survey of Halluincation in Natural Language Generation https://arxiv.org/abs/2202.03629

**Language requirements:** English

**Supervisor name and email:** Alexander Ilin (alexander.ilin@aalto.fi) and Sam Spilsbury (sam.spilsbury@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:** Alexander Ilin

**Topic available:** yes_one_instance

**Topic available also for a group:** yes (max 2 students in the group)

---

**Topic #21: Machine learning inference at network edge systems**

**Background:** Many modern systems are based on wireless connectivity and sensor data processing [1]. The list of examples is almost newer ending: autonomous cars, smart packet logistics, intelligent light systems, augmented game systems, smart homes, health systems, sport systems, etc. From the engineering point of view, complexity is often the primary challenge, and therefore, applying modern machine learning has emerged as an realistic way to tackle the challenges.

Mobility and distribution of the related computing services are inherent for wireless systems processing sensor data [2]. In order to manage the computation, the processing requirements of the various inference tasks need to be estimated. Also, the inference machinery needs to be optimized for the specific purposes. [3]

**Task:** Make a review of state-of-the-art inference systems and study the operation of typical practical inference systems by using CARLA [4] and PyTorch [5]. (If you check CARLA and its demo videos and imagine a distributed computing system with inference services there, you probably get rather well an understanding what this topic is about.)

**Prerequisites:** Basics of deep learning (CS-E4890 Deep Learning), good programming skills (and Python/PyTorch knowledge) and understanding of systems (e.g., some systems course like CS-C3140) are the primary prerequisites. But also, interest in/understanding of game engines (CARLA is based on one) and knowledge on distributed systems is useful.

**References:**

[1] P. Pop, M. Törngren (eds.). Nordic Industrial IoT Roadmap. https://www.nordic-iot.org/roadmap/
[2] Yousefpour & al. All One Needs to Know about Fog Computing and Related Edge Computing Paradigms - A Complete Survey. Journal of Systems Architecture. DOI:10.1016/j.sysarc.2019.02.009
[3] Zhuoqing Chang, Shubo Liu, Xingxing Xiong, Zhaohui Cai, and Guoqing Tu. A survey of recent advances in edge-computing-powered articial intelligence of things. IEEE Internet of Things Journal, 8(18):13849–13875, 2021. DOI:10.1109/JIOT.2021.3088875
[4] https://carla.org/
[5] https://pytorch.org/

**Language requirements:** English

**Supervisor name and email:** Vesa Hirvisalo (firstname.lastname@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:** Vesa Hirvisalo

**Topic available:** yes_one_instance

**Topic available also for a group:** yes (max 2 students in the group)

---

**Topic #22: Multi-agent reinforcement learning**

**Background:** Reinforcement Learning (RL) is a branch of Machine Learning, where agents learn to optimize an external reward signal by trial-and-error. Due to sample-inefficiency and the trial-and-error nature of RL, RL is typically used in combination with a real-time simulator. Therefore, game engines and games provide a natural platform for researching new RL algorithms [0].

In Multi-Agent RL (MARL) [1] agents learn to collaborate and communicate with each other to achieve a common goal. For example, MARL agents can learn to compete in a 3v3 soccer game or to collaborate when moving heavy furniture.

The task is to review MARL algorithms and create a hands-on experiment Unity [2] and Ray [3].

**Prerequisites:** Basics of deep learning (CS-E4890 Deep Learning), programming skills (Python, C#), experience or interest towards using game engines (Unity)

**References:**

[0] Mnih, V. *et al.* (2015) ‘Human-level control through deep reinforcement learning’, doi:10.1038/nature14236

[1] Hernandez-Leal, P., Kartal, B. and Taylor, M.E. (2018) ‘A Survey and Critique of Multiagent Deep Reinforcement Learning’, http://arxiv.org/abs/1810.05587

[2] https://unity.com/

[3] https://docs.ray.io/en/latest/index.html

**Language requirements:** English

**Supervisor name and email:** Anton Debner (firstname.lastname@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:** Vesa Hirvisalo

**Topic available:** yes_one_instance

**Topic available also for a group:** yes (max 2 students in the group)

---

**Topic #23: Diffusion Autoencoder with Structured Latent Representations**

**Background:** Diffusion-based models have recently become popular for high quality conditional and unconditional image generation. Conditional diffusion models have been shown to work with both latent image representations [1] as well as text embeddings [2] as the conditioning signal. On the other hand, the self-supervised object learning models such as Slot Attention [3, 4] take images as input and separate individual objects from the scene into compact representations stored in structured latent “slots”, which are then used to reconstruct the original image. The aim of this project is to combine the ideas from the condifional diffusion-based autoencoders with those from object-based learning and perform experiments for image autoencoding with structured object representation methods based on slot attention or similar method applied to simple image data.

**Prerequisites:** Basics of Deep Learning (CS-E4890 Deep Learning), or comparable knowledge/experience and Python programming skills. Experience with PyTorch and Python scientific programming a strong plus.

**References:**

[1] Preechakul et al., 2022, 'Diffusion Autoencoders: Toward a Meaningful and Decodable Representation', https://arxiv.org/abs/2111.15640
[2] Ramesh et al., 2022, 'Hierarchical Text-Conditional Image Generation with CLIP Latents', https://arxiv.org/abs/2204.06125
[3] Locatello et al., 2020, 'Object-Centric Learning with Slot Attention', https://arxiv.org/abs/2006.15055
[4] Singh et al., 2022, 'Illiterate DALL-E Learns to Compose', https://arxiv.org/abs/2110.11405

**Language requirements:** English

**Supervisor name and email:** Alexander Ilin (alexander.ilin@aalto.fi) and/or Arturs Polis (arturs.polis@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:** Alexander Ilin

**Topic available:** yes_one_instance

**Topic available also for a group:** yes (max 2 students in the group)

---

**Topic 24: Central composite design integration**

**Background:** Central composite design integration is a deterministic quadrature integration method useful for low dimensional posteriors with costly log posterior density evaluations. Typical use case is when the latent variables in a latent Gaussian variable model are integrated out using Laplace's method. If the number of the latent values is high, running the Laplace's method makes the computation of the lower dimensional marginal posterior slow. However, this marginal posterior has often nice shape, and central composite design integration can provide reasonable accuracy with much less evaluations than, e. g. MCMC. The project involves implementing CCD algorithm and diagnostics in R.

**Prerequisites:** Bayesian inference and Monte Carlo. R skills.

**Language requirements:** English

**Supervisor name and email:** Aki Vehtari (aki.vehtari@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:**

**Topic available:** yes_one_instance

**Topic available also for a group:** yes

---

**Topic #25: Semi-Supervised Video Action Recognition**

**Background:** Semi-supervised learning seeks to learn from both unlabeled and labeled data sampled from the same distribution. Recently, semi-supervised models for image recognition have performed comparably to, sometimes even better than, its supervised counterparts. However, semi-supervised video action recognition has received very little attention. In this project, we will investigate recent semi-supervised approaches. We will extend two of them on top of vision transformer (ViT) backbones for video understanding and compare their performances for action recognition on benchmark datasets.

Few references:

Ankit Singh, et al. Semi-Supervised Action Recognition with Temporal Contrastive Learning, https://arxiv.org/pdf/2102.02751.pdf

Kihyuk Sohn, et al. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence,https://arxiv.org/pdf/2001.07685.pdf

**Prerequisites:** Basics of deep learning, Python and PyTorch knowledge

**Language requirements:** English

**Supervisor name and email:** Selen Pehlivan (selen.pehlivantort@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:** Jorma Laaksonen

**Topic available:** yes_one_instance

**Topic available also for a group:** yes (max 2 students in the group)

---

**Topic #26: HMC for Bayesian ODEs with Gaussian processes**

**Background:** Recently a method has been proposed to perform variational inference
for unknown ODEs using Gaussian processes (see https://arxiv.org/abs/2106.10905),
enabling approximate Bayesian inference for unknown ODEs within minutes.
We have recently implemented a prototype which follows a similar method,
but appears to enable us to perform "full" (generally much more expensive)
Bayesian inference via HMC within seconds. The project involves a review of the
method and of its predecessors as well as experiments to compare efficiency and accuracy.

**Prerequisites:** Bayesian inference and MCMC. Python+Stan knowledge ideal, but Python+X, R+Stan or Julia+X work as well. Some knowledge of ODEs and GPs helpful.

**Language requirements:** English

**Supervisor name and email:** Nikolas Siccha (nikolas.siccha@aalto.fi)

**Supervisor's research group leader's name unless supervisor him/herself is one:** Aki Vehtari

**Topic available:** yes_one_instance

**Topic available also for a group:** no

---

**Topic #27: Multilabel Classification for Music and Audio using Deep Learning**

**Background**:

In this project, you will work with audio and music data to develop multi-label classification systems that are robust to acoustically challenging conditions. Some applications include audio tagging, sound event localization and detection, music genre classification, and other audio tasks. These tasks are challenging due to the large size of data, and the non-linear human perception of sound which is influenced by psycho-acoustics.

The goal of the project is to explore different deep learning architectures (CNNs, RNNs, Wavenet, etc), as well as data augmentation techniques, to improve the performance of specific tasks for acoustically challenging scenarios. These scenarios include poor quality loudspeakers (e.g. mobile phones) or highly reverberant rooms. Ideally, the solutions developed will beside to participate in competitions such as DCASE, MediaEval, or MIREX.

This is a vast project with many possible tasks for enthusiastic students. Some activities could include literature search, coding of different parts (pre-processing, signal processing, deep learning, evaluation, auralization, acoustic simulations), or running experiments and collecting results.

Related materials:

- Purwins, H., B. Li, T. Virtanen, J. Schlüter, S. Chang, and T. Sainath. “Deep Learning for Audio Signal Processing.” https://doi.org/10.1109/JSTSP.2019.2908700
- A. Mesaros, T. Heittola, T. Virtanen and M. D. Plumbley, "Sound Event Detection: A tutorial," in *IEEE Signal Processing Magazine*, vol. 38, no. 5, pp. 67-83, Sept. 2021, doi: https://doig.org/10.1109/MSP.2021.3090678

- Nam, J., K. Choi, J. Lee, S. Chou, and Y. Yang. “Deep Learning for Audio-Based Music Classification and Tagging: Teaching Computers to Distinguish Rock from Bach.” https://doi.org/10.1109/MSP.2018.2874383
  Choi, Keunwoo, György Fazekas, Kyunghyun Cho, and Mark Sandler. “A Tutorial on Deep Learning for Music Information Retrieval.” http://arxiv.org/abs/1709.04396.
- Pons, Jordi, Oriol Nieto, Matthew Prockup, Erik Schmidt, Andreas Ehmann, and Xavier Serra. “End-to-End Learning for Music Audio Tagging at Scale.” http://arxiv.org/abs/1711.02520.
- H. Daolang and Falcon-Perez R., “SSELDNET: A FULLY END-TO-END SAMPLE-LEVEL FRAMEWORK FOR SOUND EVENT LOCALIZATION AND DETECTION,” DCASE2021 Challenge, Nov. 2021.
- P.-A. Grumiaux, S. Kiti, L. Girin, and A. Guérin, “A review of sound source localization with deep learning methods,” arXiv [cs.SD], 2021.
- T. Kim, J. Lee and J. Nam, "Comparison and Analysis of SampleCNN Architectures for Audio Classification," in IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 2, pp. 285-297, May 2019, doi: 10.1109/JSTSP.2019.2909479. https://ieeexplore.ieee.org/document/8681654


**Prerequisites**:

- Good knowledge of deep learning, programming skills in python.
- Bonus: Experience with PyTorch.
- Bonus: Familiarity with audio signal processing concepts such as filtering, time-frequency transforms, impulse responses, sampling frequency, pitch, etc ...


**Supervisor**: Ricardo, Falcon Perez (ricardo.falconperez@aalto.fi)


**Supervisor's research group leader's name unless supervisor him/herself is one:** Ville Pulkki

**Topic available:** yes_many_instances

**Topic available also for a group:** yes **(**max 2)