MLOps

Model Development & Offline Evaluation

“Designing Machine Learning Systems

- Introduction to ML system design
- Data Engineering fundamentals
- Training data & Feature Engineering
- Model development and offline evaluation
- Model deployment & prediction services
- Introduction to MLOps

“Model Development & Offline Evaluation

Model Development & Offline Evaluation

Intended Learning Outcomes (ILOs)

- After this lecture, you should be able to:
  - Explain the iterative process of ML model development and offline evaluation.
  - Compare and contrast different ML algorithms and their suitability for various tasks.
  - Evaluate ML models based on multiple criteria, including performance, simplicity, and generalization.
  - Analyze trade-offs in model selection, considering both current and potential future performance.
  - Identify and assess key assumptions underlying different ML models and their implications for model performance and reliability.

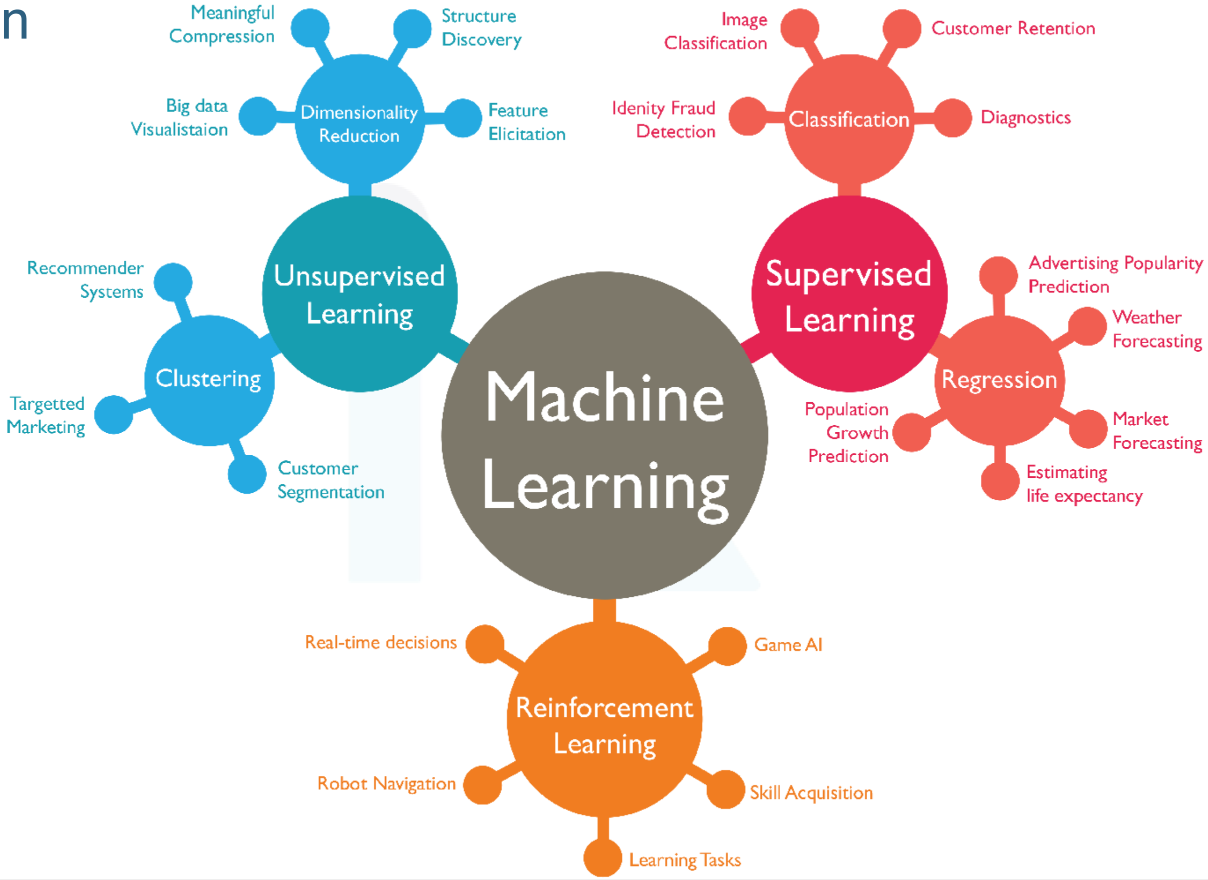
Model Development & Offline Evaluation

Introduction

- ML model development is an iterative process
  - Compare the model’s performance against its performance in previous iterations
  - Use perturbation tests, invariance tests, model calibration and slide-based tests to evaluate the ML model before deployment

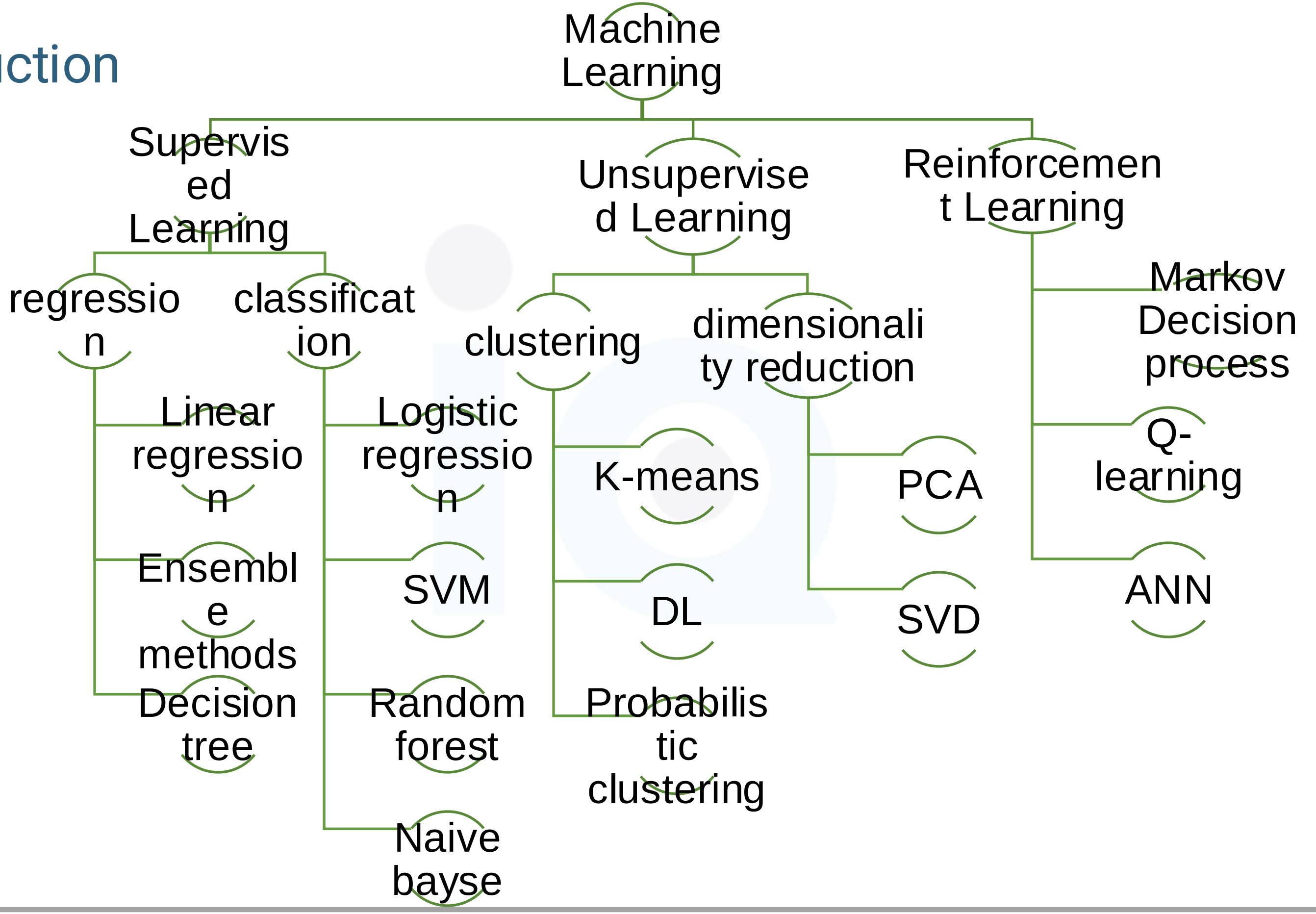
Model Development & Offline Evaluation

Introduction



Model Development & Offline Evaluation

Introduction



Model Development & Offline Evaluation

Veracity Check

1. True or False: ML model development is a one-time process that doesn't require iteration.
2. True or False: Perturbation tests and invariance tests are used to evaluate ML models before deployment.
3. True or False: Supervised learning includes only regression and classification tasks.
4. True or False: Deep Learning is the only useful ML algorithm for production environments.

Model Development & Offline Evaluation

Veracity Check

1. True or False: ML model development is a one-time process that doesn't require iteration.

   Answer: False Justification: ML model development is explicitly described as an iterative process, where the model's performance is compared against previous iterations.

2. True or False: Perturbation tests and invariance tests are used to evaluate ML models before deployment.

   Answer: True Justification: The lecture mentions using perturbation tests, invariance tests, model calibration, and slide-based tests to evaluate ML models before deployment.

3. True or False: Supervised learning includes only regression and classification tasks.

   Answer: False Justification: While supervised learning does include regression and classification, the lecture also mentions other types of machine learning such as unsupervised learning and reinforcement learning.

4. True or False: Deep Learning is the only useful ML algorithm for production environments.

   Answer: False Justification: The lecture explicitly states that even though Deep Learning is finding more use cases in production, other ML algorithms are still very useful.

“Evaluating ML Models

Model Development & Offline Evaluation

ML Model Selection

- Given a task, what ML algorithm should you use for it?
  - Colleague mentioned that Gradient-boosted trees always worked for it.
  - Can you select a model based on anecdotal evidences?
  - Note! Time and compute resources are limited
  - Neural Network, in particular, Deep Learning (DL) is not the only ML algorithm
  - Even though DL is finding more use cases in production, other ML algorithms are very useful
  - Hybrid ML models, combining pretrained NN (like BERT or GPT) and logistic regression, might be more powerful

Model Development & Offline Evaluation

ML Model Selection

- Want to deduct fraudulent transaction?
  - It is an abnormality problem
  - Use E.g.; k-nearest neighbors, isolation forest, clustering and NN
  - Non-NN algorithms are tend to to more explainable
- Consider not only accuracy, F1 score and log loss but also
  - How much data, compute, training time, inference latency, interpretability, etc.
- Fast moving
  - No matter how the algorithm is accurate today, it will be replaced soon
    - LSTM-RNNs (popular in 2016) was replaced by transformer-based architectures for NLP tasks
    - Check NeurIPS, ICLR and ICML conferences

Model Development & Offline Evaluation

Do the Foundational Models Replace ML?

Model Development & Offline Evaluation

Veracity Check

1. True or False: It's always best to use the most complex ML algorithm available for a given task.
2. True or False: Anecdotal evidence is a reliable basis for selecting an ML model.
3. True or False: Non-neural network algorithms tend to be more explainable than deep learning models.
4. True or False: The best ML model for a task will remain the same over time.

Model Development & Offline Evaluation

Veracity Check

1. True or False: It's always best to use the most complex ML algorithm available for a given task.

   Answer: False Justification: The lecture emphasizes starting with simpler models and considering factors beyond just accuracy, such as interpretability, computational requirements, and available data.

2. True or False: Anecdotal evidence is a reliable basis for selecting an ML model.

   Answer: False Justification: The lecture cautions against selecting models based on anecdotal evidence, suggesting a more systematic approach to model selection.

3. True or False: Non-neural network algorithms tend to be more explainable than deep learning models.

   Answer: True Justification: The lecture mentions that non-NN algorithms tend to be more explainable, which is an important consideration in model selection.

4. True or False: The best ML model for a task will remain the same over time.

   Answer: False Justification: The lecture describes ML as fast-moving, stating that no matter how accurate an algorithm is today, it will likely be replaced soon, giving the example of LSTM-RNNs being replaced by transformer-based architectures for NLP tasks.

Model Selection

Avoid the Star-of-the-Art Trap (1/6)

- Understand the problem
  - Prioritize problem understanding over chasing the latest models
  - Ensure a clear understanding of the problem's requirements, constraints, and available data before selecting a ML model
- Evaluate simplicity
  - Complex models often come with higher computational requirements, increased risk of overfitting, and reduced interpretability
  - Consider simpler models that can provide good performance while being easier to understand and maintain

Model Selection

Avoid the Star-of-the-Art Trap (1/6)

- Prioritize generalization
  - While state-of-the-art models may excel on specific benchmark datasets, they may not generalize well to real-world scenarios
  - Prioritize models with strong generalization capabilities to ensure reliable performance in various situations
- Consider data requirements
  - Some models may have specific data requirements, such as large-scale labeled datasets or extensive preprocessing.
  - Assess the availability and feasibility of acquiring the necessary data for training and deploying the chosen model

Model Selection

Avoid the Star-of-the-Art Trap (1/6)

- Benchmark against baseline models
  - Compare the performance of cutting-edge models against established baseline models that are well-understood and widely used
  - Allows a fair assessment of improvements and helps avoid the hype surrounding new models
- Incorporate ensemble methods
  - Combine multiple models to leverage their strengths and compensate for their weaknesses
  - Provide robust predictions and reduce the risk of relying solely on a single
  - Remember, selecting the most suitable ML model involves considering multiple factors beyond just the latest advancements
  - Prioritize practicality, interpretability, and generalization to make informed decisions and avoid falling into the star-of-the-art trap

Model Selection

Start with the simplest models (2/6)

- Occam's Razor
  - The simplest solution is often the best one. Apply this principle to ML model selection by starting with simpler models before considering more complex ones
- Simplicity promotes understanding
  - Simple models, such as linear regression or decision trees, are easier to understand and interpret. They provide insights into the relationships between variables
- Reduced computational requirements
  - Require fewer computational resources, making them faster to train and deploy. This can be particularly beneficial when working with large datasets or resource-constrained environments
- Avoid overfitting

Model Selection

Start with the simplest models (2/6)

- Iterative improvement
  - Allows for incremental improvement and allows to assess the model's performance.
  - Incrementally introduce complexity, only if necessary
- Robustness to noise
  - Often more robust to noise and outliers in the data, and less likely to be influenced by irrelevant or noisy features, resulting in more stable and reliable predictions
- Rapid prototyping
  - Enables quick prototyping and experimentation, and allows to build and test initial models rapidly, allowing for iterative development and refinement of ML solution.
- Benchmark for performance comparison

Model Selection

Avoid human biases in selecting models (3/6)

- Define objective criteria
  - Establish clear and objective criteria for selecting ML models. This can include factors like accuracy, interpretability, computational requirements, and scalability
- Blind evaluation
  - Conduct blind evaluations where the identities of the model creators or their affiliations are concealed. This helps prevent biases based on reputation or personal relationships
- Diverse evaluation team
  - Form a diverse evaluation team comprising individuals with different backgrounds, perspectives, and expertise. This can help mitigate unconscious biases

Model Selection

Avoid human biases in selecting models (3/6)

- Double-blind peer review
  - Implement a double-blind peer review process, where both the reviewers and model creators are anonymous to each other. This prevents biases related to gender, ethnicity, etc.
- Consider counterfactuals
  - Explore counterfactual scenarios by considering alternative models or approaches. This can help challenge preconceived notions and biases, encouraging a more comprehensive and unbiased evaluation process
- Use standardized benchmarks
  - Rely on standardized benchmarks and datasets to compare and evaluate models objectively. These benchmarks provide a level playing field for different models and reduce biases that may arise from using biased or unrepresentative data

Model Selection

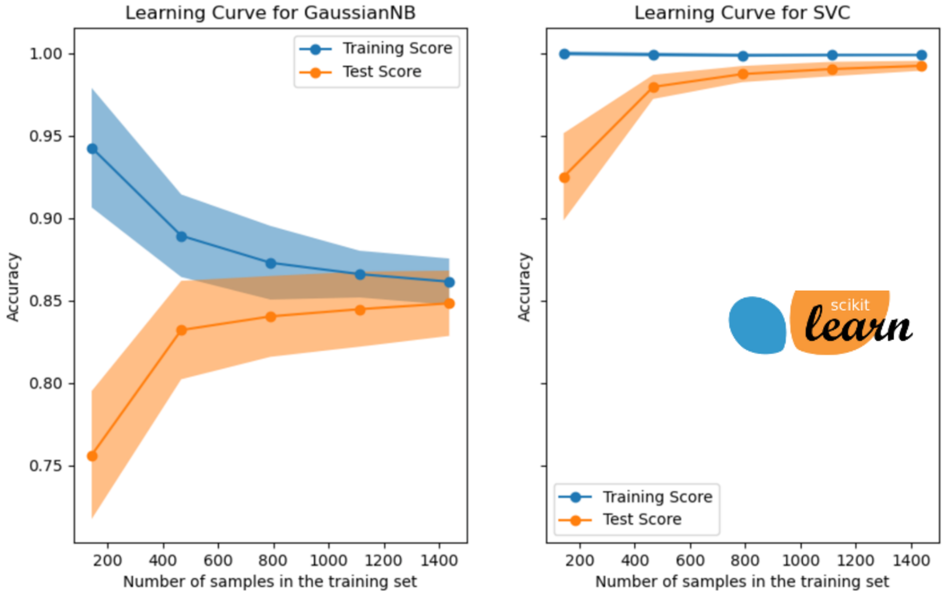
Avoid human biases in selecting models (3/6)

- Document the decision-making process
  - Keep a record of the decision-making process, including the criteria used, evaluation results, and the reasoning behind selecting or rejecting specific models
  - This promotes transparency and allows for review and scrutiny of the decision-making process
- Regularly reassess models
  - Continuously reassess selected models to ensure they are still the best fit for the problem at hand
  - New developments and advancements may require reevaluation and adjustment of the chosen models to avoid being anchored to previous biases

Model Selection

Evaluate good performance now Vs. better performance later (4/6)

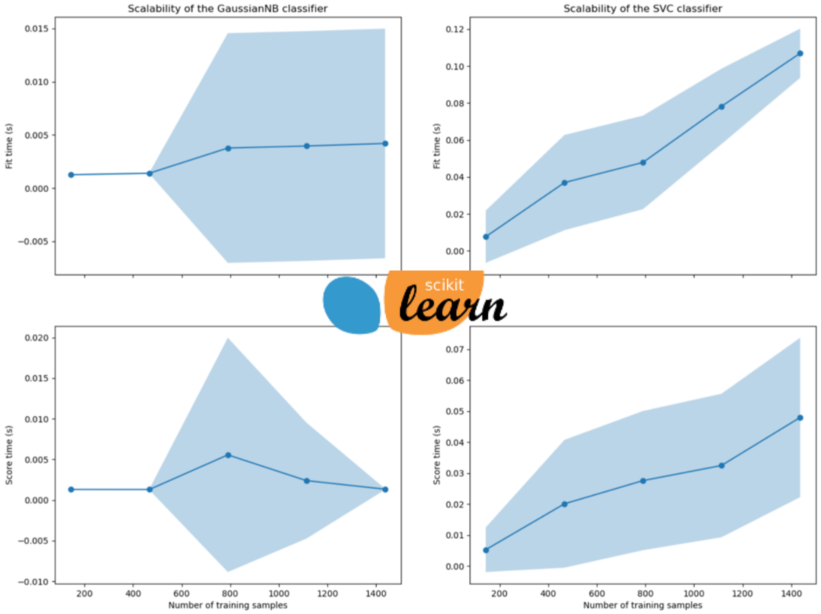
- Model evolution
  - The best model now does not always mean the best model a few months from now
  - A tree-based model might work better now with less training data, but a few months later a NN model might outperform when more training data is available
  - Use learning curve to estimate evaluate the change of the model performance
    - Need not provide exact performance gain but indicate



Model Selection

Evaluate good performance now Vs. better performance later (4/6)

- Complexity analysis
  - Scalability of the models in terms of training and scoring times also needs to be considered



- SVM classifier complexity at fit and score time increases rapidly with the number of samples (more than quadratic)
- In contrast, the naive Bayes classifier scales much better with a lower complexity at fit and score time.

Model Selection

Evaluate trade-offs (5/6)

- Define the trade-offs
  - Identify the different factors or variables involved in the decision-making process and understand the trade-offs between them
  - This may include performance vs. cost, accuracy vs. speed, complexity vs. interpretability, etc.
- Prioritize objectives
  - Determine the most important objectives based on the specific context and requirements.
  - Clearly define the goals and desired outcomes to guide the evaluation process and make informed trade-offs
- Quantify metrics
  - Establish measurable metrics to evaluate the trade-offs. Use quantitative measures such as accuracy, precision, recall, processing time, cost, and resource utilization to compare different options objectively

Model Selection

Evaluate trade-offs (5/6)

- Assess risks
  - Consider the potential risks associated with each option. Evaluate factors such as model robustness, potential biases, ethical considerations, regulatory compliance, and security implications to make well-informed decisions
- Understand constraints
  - Identify any limitations or constraints that may impact the decision-making process
  - This could include computational resources, time constraints, available data, expertise, or budget limitations
- Involve stakeholders
  - Seek input from relevant stakeholders who will be impacted by the decision
  - Collaborate with domain experts, end-users, and decision-makers to ensure a holistic understanding of the trade-offs and gather diverse perspectives

Model Selection

Evaluate trade-offs (5/6)

- Perform sensitivity analysis
  - Conduct sensitivity analysis to understand the impact of varying the trade-offs. Evaluate how different factors affect the overall performance and outcomes, allowing for a more comprehensive decision-making process
- Consider long-term implications
  - Look beyond short-term gains and consider the long-term implications of the trade-offs
  - Assess factors such as scalability, maintainability, adaptability, and future-proofing to make decisions that align with long-term goals
- Document decision rationale
  - Keep a record of the decision-making process, including the trade-offs considered, the metrics used, and the rationale behind the final decision.
  - This documentation can aid in future evaluations and provide transparency.

Model Selection

Understand the model assumptions (6/6)

- Model assumptions
  - Every ML model is built upon certain assumptions about the data and the relationship between variables
  - Understanding these assumptions is crucial for accurate interpretation and reliable predictions
- Linearity assumption
  - Linear models, such as linear regression, assume a linear relationship between the input features and the target variable
  - Assess whether this assumption holds in the context of the problem at hand
- Independence assumption
  - Many models assume that the observations are independent of each other
  - Violating this assumption, such as in time series or spatial data, can lead to biased or inefficient results. Consider models that account for dependencies

Model Selection

Understand the model assumptions (6/6)

- Normality assumption
  - Some models, like linear regression, assume that the residuals (the differences between predicted and actual values) follow a normal distribution
  - Assess whether the data exhibits normality and consider alternative models if it does not
- Homoscedasticity assumption
  - Homoscedasticity refers to the assumption that the variability of the residuals is constant across all levels of the independent variables
  - Violations of this assumption may indicate the need for more advanced modeling techniques
- Feature independence assumption
  - Certain models assume that the input features are independent of each other
  - If there are strong correlations or dependencies among the features, consider

Model Selection

Understand the model assumptions (6/6)

- Stationarity assumption
  - Time series models often assume the statistical properties of the data remain constant over time
  - Assess whether the data exhibits stationary behavior or if transformations or additional techniques are needed
- Data distribution assumption
  - Different models have different distributional assumptions for the input data
  - For example, some models assume the data is normally distributed, while others work well with non-normal or categorical data
- Heterogeneity assumption
  - Some models assume homogeneity of the data, meaning that the relationships between variables are consistent across all subgroups. Consider models that allow for heterogeneity if the relationships vary significantly across different groups

Designing Machine Learning Systems

Summary

- Introduction to ML system design
- Data Engineering fundamentals
- Training data & Feature Engineering
- Model development and offline evaluation
- Model deployment & prediction services (Next Module)
- Introduction to MLOps (Next Module)