Supervised Learning

CSIP5403 – Research Methods and AI Applications

Dr. Nathanael L. Baisa

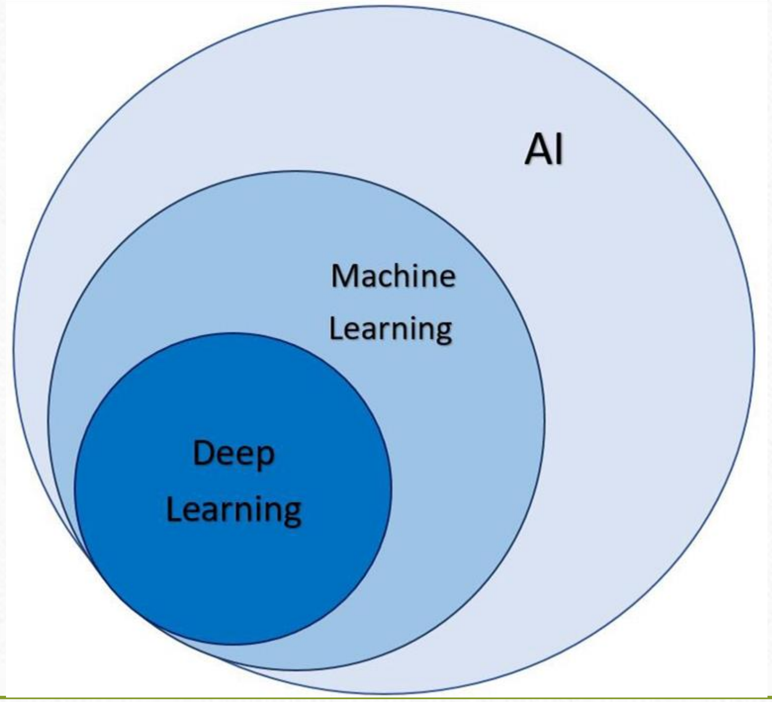
Contents

Session Outcomes

- Gain awareness of different ML branches.
- Acquire proper understanding of supervised learning methods.
- Understand how to apply supervised learning methods for different applications and how to properly evaluate their performance.

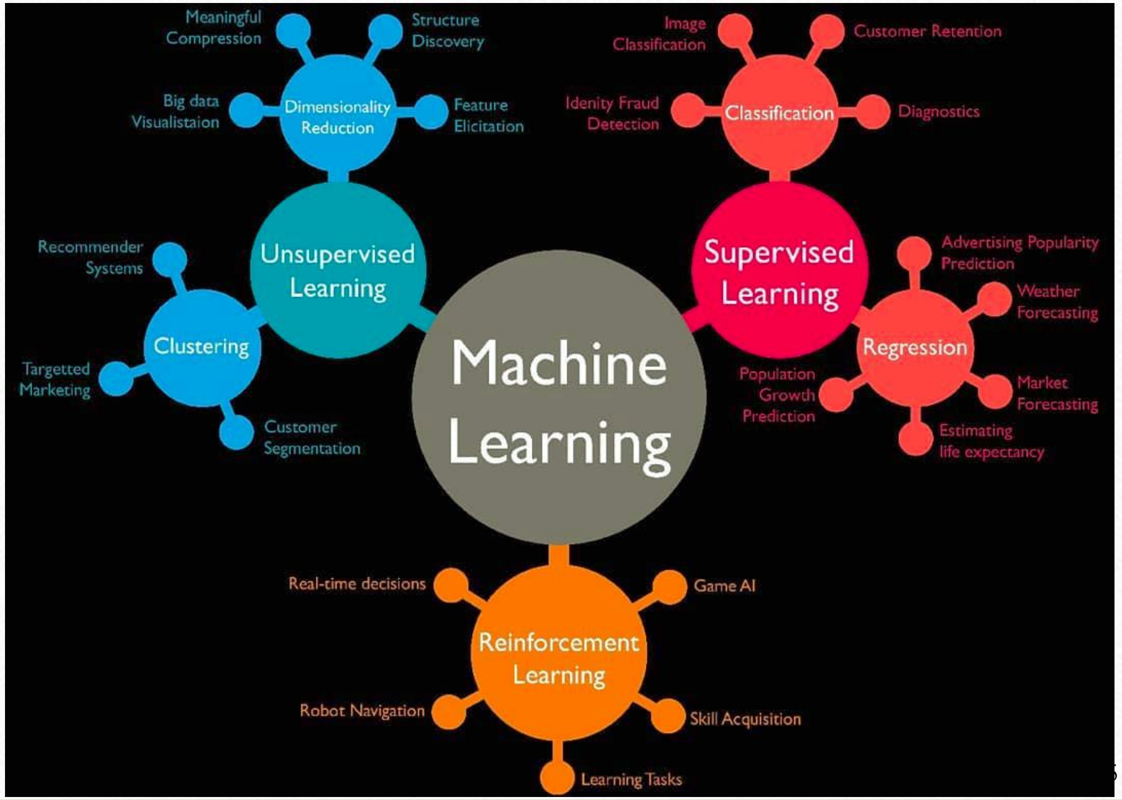
1. Introduction

- ML is a branch of AI (a field concerned with designing intelligent systems).
- ML algorithms build models from training data to make predictions, without being explicitly programmed to do so.
- A ML algorithm has parameters (weights) whose values are learned from training data.



1. Introduction

- ML has 3 main branches for different applications:

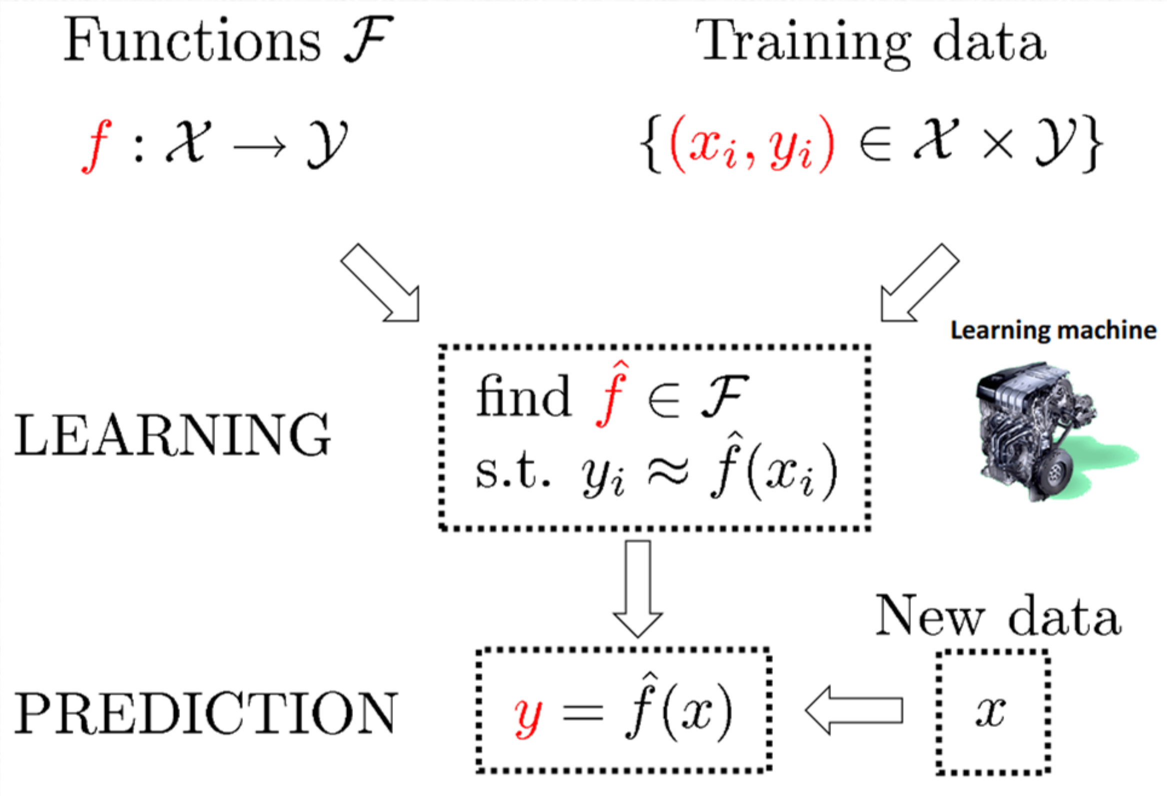


2. Supervised Learning Methods

- Dataset needs to be labelled.
- Each dataset should contain a good selection of instances to be representative of the application domain.
- Two types:
  - **Regression** - predicts continuous valued output.
    E.g. linear regression, non-linear regression, etc
  - **Classification** - predicts discrete valued output.
    e.g. logistic regression, decision trees and ensemble methods, SVM, Neural Networks (Perceptron, MLP, etc.), etc.

2. Supervised Learning Methods

- Supervised learning:



2. Supervised Learning Methods

- Learning is performed by optimizing a cost function using optimization algorithms, for instance, using gradient descent.

In general Minimize with respect to $f \in \mathcal{F}$

$$\sum_{i=1}^{N} l\left(f\left(x_i\right), y_i\right) + \lambda R\left(f\right)$$

- Choose loss function for regression, classification, etc. such as MSE loss, CE loss (softmax loss), hinge loss, contrastive loss, etc.
- Choose regularization function: $R(f) = \|\mathbf{w}\|_2^2$ for L2 (Ridge)

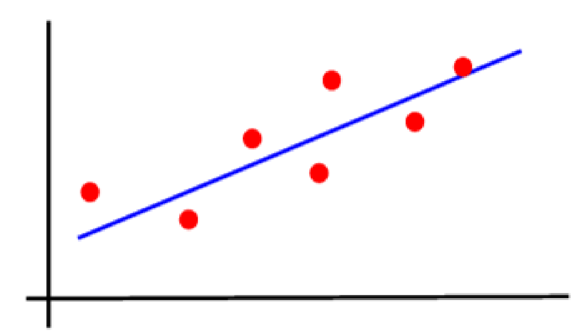
  $R(f) = \|\mathbf{w}\|_1$ for L1 (Lasso) - sparse.

2. Supervised Learning Methods

- **Regression:**
  - Given a training set of N observations $(x_1, \ldots, x_N)$ and $(y_1, \ldots, y_N)$, $x_i, y_i \in \mathbb{R}$, regression problem is to estimate f(x) from this data.

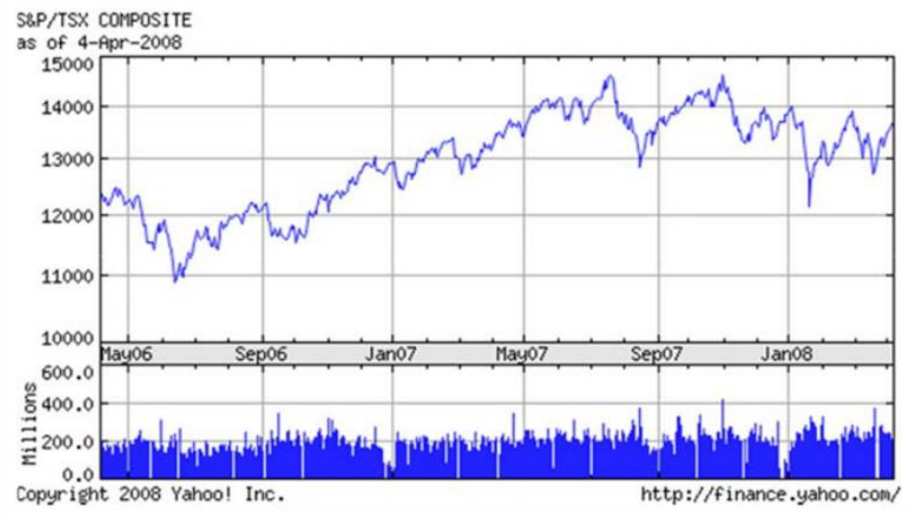
    e.g. linear regression

    $$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$$



2. Supervised Learning Methods

- **Regression:**
  - Applications include stock price prediction:



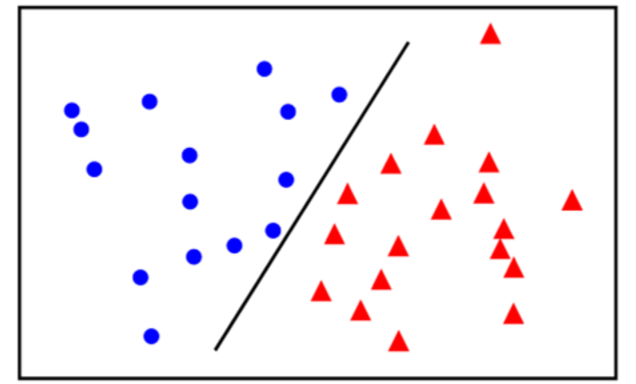
2. Supervised Learning Methods

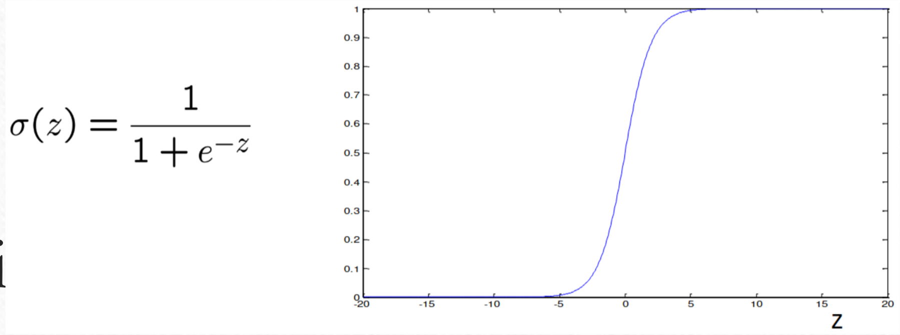
- **Classification:**
  - Given a training set of N observations $(x_1, \ldots, x_N)$ and $(y_1, \ldots, y_N)$, $x_i \in \mathbb{R}^d$, $y_i \in \{-1, 1\}$, classification problem is to estimate f(x) from this data.

    e.g. logistic regression

    $$f(\mathbf{x}) = \sigma(\mathbf{w}^T\mathbf{x} + b)$$

    σ is logistic (sigmoid) functi





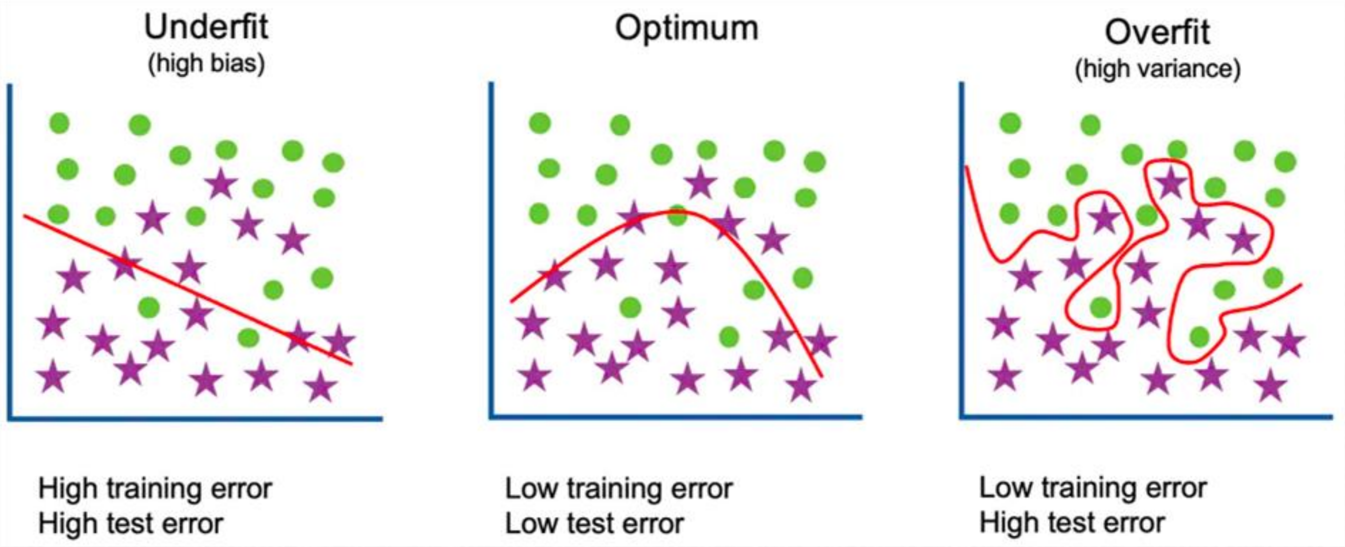
2. Supervised Learning Methods

- **Classification:**
  - Multi-class classification - assign input vector **x** into one of K classes C_K.
  - Build from binary classifiers, e.g. logistic regression (LR). Learn K two-class 1-vs-the-rest classifiers $f_K(\mathbf{x})$ and choose class with most positive score.
  - Use Softmax which generalizes LR for multiple classes.

$$P(C_k|\mathbf{x}) = \frac{\exp(f_k(\mathbf{x}))}{\sum_j^K \exp(f_j(\mathbf{x}))}$$ where $f_K(\mathbf{x}) = \mathbf{w}_K^T\mathbf{x} + b_K$

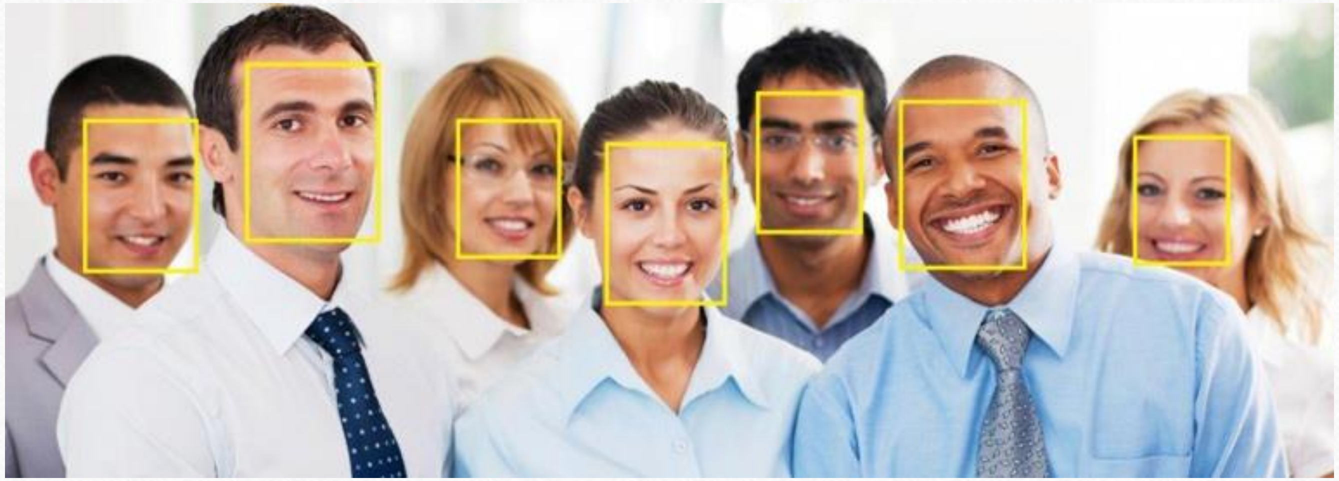
2. Supervised Learning Methods

- **Classification:**
  - The complexity of the (discriminant) function needs to be controlled to handle the generalization problem.



2. Supervised Learning Methods

- **Classification:**
  - Applications include face detection.



→ Classify an image window into face and non-face classes.

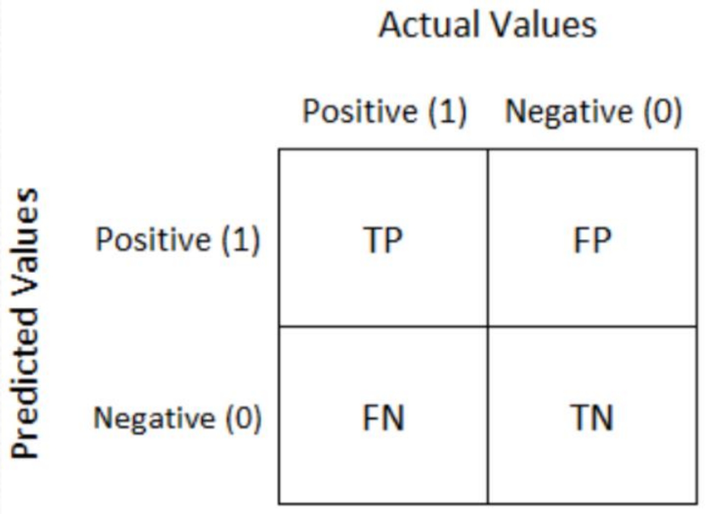
3. Evaluation Metrics

- **Evaluation Metrics for Regression:**
  - Mean Squared Error (MSE) – the lower the better.
  - Root Mean Squared Error (RMSE) – the lower the better.
  - Mean Absolute Error (MAE) – the lower the better.
  - R^2, pronounced 'R-squared', computes the coefficient of determination, which is the proportion of the variation in the dependent variable that is predictable from the independent variable(s). The higher the better.

  - Read more on: https://medium.com/analytics-vidhya/evaluation-metrics-for-regression-models-c91c65d73af

3. Evaluation Metrics

- **Evaluation Metrics for Classification:**
  - Accuracy
  - Precision
  - Recall
  - F1-score
  - Confusion matrix:



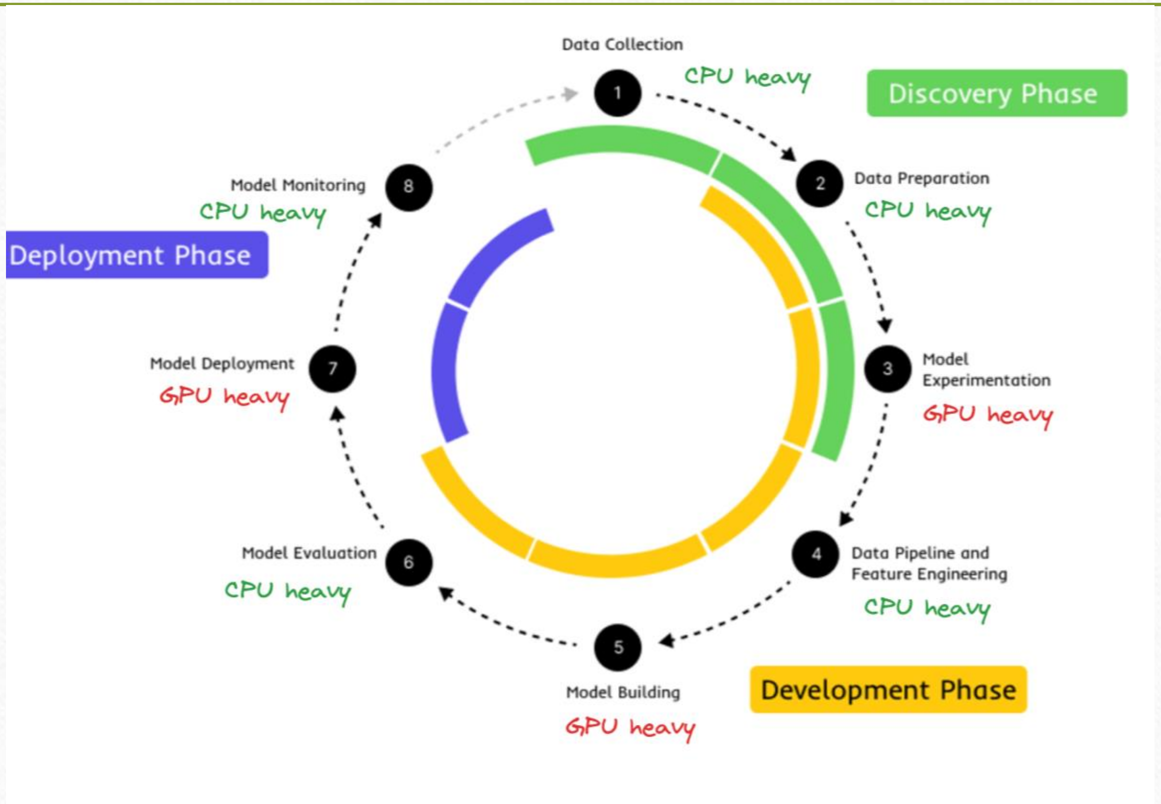
  - Read more on: https://medium.com/@impythonprogrammer/evaluation-metrics-for-classification-fc770511052d

4. Practical Tips on Supervised Learning

- Split the dataset into training, validation and test subsets.
- Choose parameters on validation set e.g. cross-validation. The validation and test sets should come from the same distribution.
- The aim of supervised learning is to do well on unseen test data.
- Handle overfitting using regularization (e.g. L1 regularization, L2 regularization) which penalizes weights with large magnitudes. Recently introduced regularization methods in deep learning include dropout, batch normalization, etc. Preprocessing of input data (normalization or standardization) is necessary in some methods.
- ML tools: Scikit-learn, PyTorch, TensorFlow, OpenCV, etc.

4. Practical Tips on Supervised Learning

- The **ML** and **Compute** Lifecycle:



5. Applications

- Regression:
  - Stock price prediction,
  - Weather forecasting, ….
- Classification:
  - Face detection,
  - Identity fraud detection, ….
- Can you list **applications** of supervised learning?
- Discuss the **advantages** and **disadvantages** of supervised learning.

6. Conclusion

- ML has 3 main branches for different applications: supervised, unsupervised and reinforcement learning.
- To effectively apply supervised learning methods to different applications, the dataset needs to be labelled and it should contain a good selection of instances to be representative of the application domain.
- It is crucial to properly handle overfitting using regularization methods.

7. References

[1]. C. Bishop, 'Pattern Recognition And Machine Learning', Springer, 2006.

[2]. K. Murphy, 'Machine Learning: A Probabilistic Perspective', MIT Press, 2012. [https://probml.github.io/pml-book/]

[3]. I. Goodfellow, Y. Bengio, A. Courville, 'Deep Learning', MIT Press, 2016.

[4]. S. Russel, P. Norvig, 'Artificial Intelligence: A Modern Approach', Pearson Series, 2021.

[5]. R. Sutton and A. Barto, 'Reinforcement Learning: An Introduction, 2nd Edition', MIT press, 2018.

[6]. R. Szeliski, 'Computer Vision: Algorithms and Applications', Springer, 2021. [https://szeliski.org/Book/]