UiT The Arctic University of Norway

Introduction to Machine Learning: Methods and Applications - Part 2

INE-3800 Operation Research I

Wahyu Andy Prastyabudi

Department of Industrial Engineering
Faculty of Engineering Science and Technology

2025-09-19

Outlines

Getting Started

Unsupervised Learning

Supervised Learning

Supplementary materials

Concluding remarks

Getting Started

Prepare the development environment

- JupyterLab hosted by UiT - https://jupyter.uit.no
- Google Colab

- JupyterLab hosted by UiT - https://jupyter.uit.no
- Google Colab

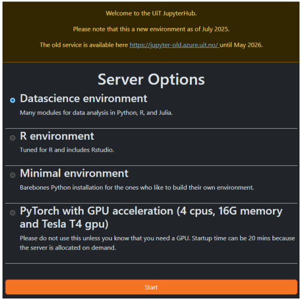

Figure 1: Select server options

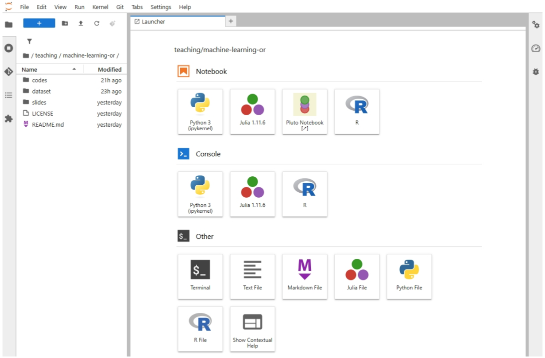

Figure 2: JupyterLab interface.

Get the resources

- Open `Terminal`
- Run the following command to download the materials

```
git clone https://github.com/wanprabu/machine-learning-or.git
```

- You may learn Python through the following channel if you are unfamiliar with Python
  - Python for Beginners - YouTube

About Jupyter Notebook

- A notebook is a shareable document that combines computer code, plain language descriptions, data, and rich visualizations.
- Read documentation for more details: https://docs.jupyter.org/en/latest/

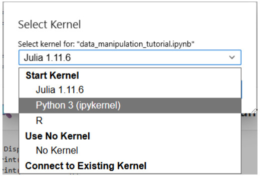

Figure 3: Select appropriate kernel.

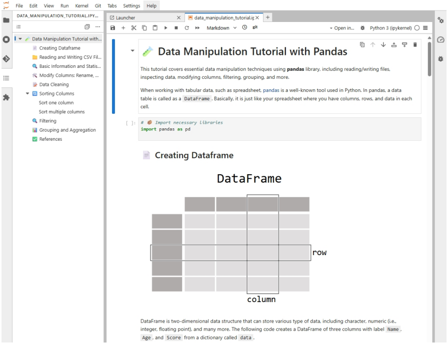

Figure 4: Jupyter notebook interface

Required libraries

- Python $>=$ 3.9
- Pandas $>=$ 2.2.0
- Numpy $>=$ 1.25.0
- Scikit-learn $>=$ 1.6.1
- Matplotlib $>=$ 3.8
- Seaborn $>=$ 0.12.2

Run the following command in `Terminal` to check the installed version

```
python --version
```

```
pip show pandas
```

Unsupervised Learning

K-means clustering

- K-means algorithm divides a set of N samples X into K disjoint clusters C
- Each described by the mean μ_j of the samples in the clusters, known as "centroids"
- Choose centroids that minimize a criterion - inertia - within cluster sum-of-squares
  - $\sum_{i=0}^{n} \min_{\mu_j \in C}(||x_i - \mu_j||^2)$
- Require to determine initial number of clusters k

Algorithm steps

- Set initial centroids by arbitrarily choosing k samples from the dataset X
- (re)assign each sample to its nearest centroid
- Create a new centroids and compute the mean value of all samples
- Repeat the last two steps until converged

K-means clustering

- K-means algorithm divides a set of N samples X into K disjoint clusters C
- Each described by the mean μ_j of the samples in the clusters, known as "centroids"
- Choose centroids that minimize a criterion - inertia - within cluster sum-of-squares
  - $\sum_{i=0}^{n} \min_{\mu_j \in C}(||x_i - \mu_j||^2)$
- Require to determine initial number of clusters k

Algorithm steps

- Set initial centroids by arbitrarily choosing k samples from the dataset X
- (re)assign each sample to its nearest centroid
- Create a new centroids and compute the mean value of all samples
- Repeat the last two steps until converged

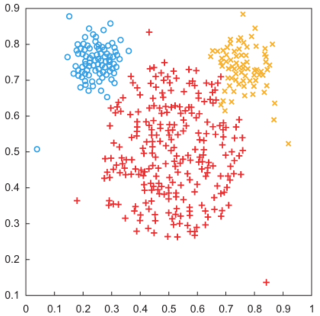

Figure 5: Example of clustering. Taken from Cielen and Meysman (2016).

- Advantages: simple and fast
- Disadvantages: dependence on input parameters, sensitive to outliers

Supervised Learning

Regression

- Regression model predicts continuous numeric values from one or more independent variables
- Regression analysis function

$$y = \theta_0 + \theta_1 x_1 + \varepsilon$$

- The goal of regression is to find a function f that maps inputs $x \in \mathbb{R}^D$ to corresponding function values $f(x) \in \mathbb{R}$, where $\mathbb{R}^D$ is vector space of real numbers
- Regression model can be formalized as

$$y_n = f(x_n, \theta) + \varepsilon$$

- y_n: the dependent variable; the thing we are trying to predict.
  - outcome variable, or response variable
- x_n: the independent variables; the input features used to train model y.
  - predictor variable, or explanatory variable
- θ: the parameter of our regression model.
  - coefficients, or weights
  - Some books use other notations (e.g., α, β)
- ε: the irreducible error.
  - A random variable that describes measurement noise, unmodeled parts of our data.

Regression (cont.)

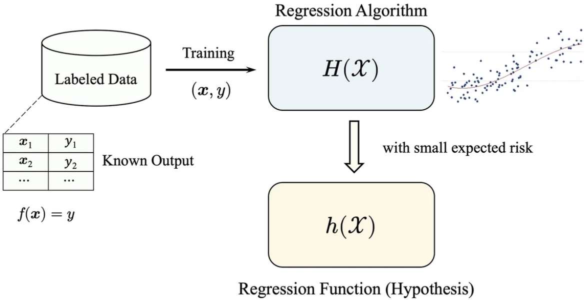

Figure 6: Training process of regression model. Taken from Wang (2025)

- Risk means minimising the errors or residuals
  - $\varepsilon_i = y_i - \hat{y}_i$
  - y_i is the actual observed value
  - $\hat{y}_i$ is the predicted value of x_i

- See the following link for visual explanation of Linear Regression

Types of regression

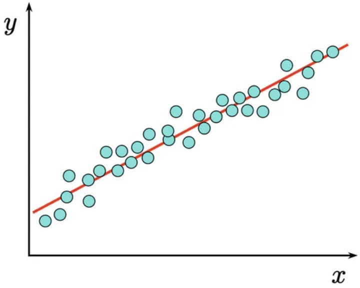

Figure 7: Linear regression illustration. Taken from Wang (2025)

- Linear regression assumes there is colinear relationship between dependent variable y and independent variable x.
- To find best fit regression line, parameter estimation is calculated using least squares, which minimize the sum of squared residuals.

- Multiple linear regression uses multiple independent variables to predict y
  - Given n training sample $S = \{(x_i, y_i) \mid i = 1, \dots, n\}$
  - For the regression model $y_i = f(x_i, \theta) + \varepsilon_i$
  - Let $x_i = (x_{i1}, \dots, x_{im})$ be multi-independent variable
  - Let $\theta = (\theta_0, \theta_1, \dots, \theta_m)$ be parameter
  - The objective function of multiple linear regression can be expressed as:

$$
\begin{aligned}
y_i &= f(x_i, \theta) + \varepsilon_i \\
&= \theta_0 + \theta_1 x_{i1} + \cdots + \theta_m x_{im} + \varepsilon_i \\
&= \sum_{j=0}^{m} \theta_j x_{ij} + \varepsilon_i,
\end{aligned}
$$

Types of regression (cont.)

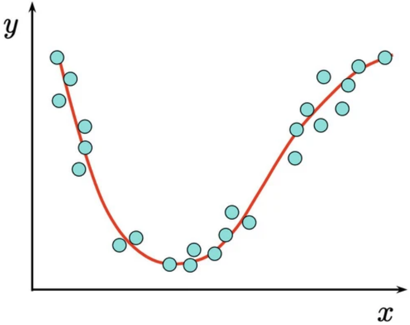

Figure 8: Polynomial regression. Taken from Wang (2025)

- Polynomial regression model assumes the relationship between the independent variable x and the dependent variable y is modelled as a p-th degree polynomial in x.
- The objective function of polynomial regression is expressed as

$$\begin{aligned} y_i &= f(x_i, \theta) + \varepsilon_i \\ &= \theta_0 + \theta_1 x_i + \theta_2 x_i^2 + \cdots + \theta_m x_i^m + \varepsilon_i \\ &= \sum_{j=0}^{m} \theta_j x_i^j + \varepsilon_i, \end{aligned}$$

- m is the degree of polynomial
- Parameter estimation is obtained by minimizing residual sum of squares

Types of regression (cont.)

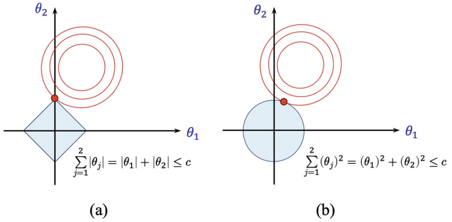

Figure 9: Lasso (a) and ridge (b) regression. Taken from Wang (2025)

- *multicollinearity* can lead to unreliable parameter estimates and high variance.
- *multicollinearity* occurs when one independent variable can be linearly predicted from the others.

- Ridge regression is a method for estimating the parameters of a linear regression model when the independent variables are highly correlated.
- Ridge regression adds a regularization term, also known as a shrinkage penalty term.
- Lasso is an abbreviation for "least absolute shrinkage and selection operator."
- Regularization of lasso regression has the effect of forcing a parameter estimate to be zero.
- Regularization of ridge regression can only make it approach zero, but cannot be equal to zero.

Classification

- Classification aims to categorize items into specific predefined class.
- ML model trains a classifier with the labelled samples and then uses it to identify which categories the other data belongs to.
- Let $\mathcal{X} \subseteq \mathbb{R}^m$ be the input space and $\mathcal{C} = \{c_j \mid j = 1, \ldots, k\}$ be the output space, consisting of k predefined class or categories.
- Each data $x_i (i = 1, \ldots, n)$ is labelled with its target class $c_j (j = 1, \ldots, k)$ based on probability distribution $P(x, c) = P(c|x) P(x)$
- The following notation maps a set of hypothesis function $H : \mathcal{X} \to \mathcal{C}$.
- The goal is to obtain $h \in H$ through the training on S such that $h(x) = \hat{c}$ has the smallest expected error.
- The target function of $f(x) = c$ can be expressed as: $\arg\min_{h \in H} R(h) = \arg\min_{h \in H} \mathbb{E}[\mathcal{L}(h(x), f(x))] = \arg\min_{h \in H} \mathbb{E}[\mathcal{L}(\hat{c}, c)]$

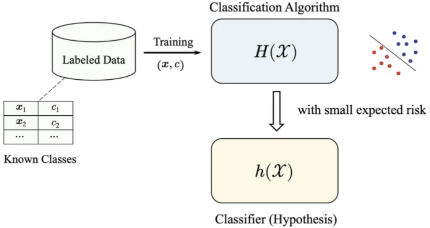

Figure 10: Training process of classification model. Taken from Wang (2025).

- $\hat{c}$ and c denote predicted and target class, respectively, whereas $\mathcal{L}$ represents the loss function

Types of classification

	Types
Functions	Probabilistic classification (e.g., logistic regression, naive Bayes)
	Statistical classification (e.g., KNN, SVM)
Tasks	one-class
	two-class
	multi-class
	imbalanced classification

Classification techniques

- Linear Models
- Support Vector Machines
- Stochastic Gradient Descent
- Nearest Neighbors
- Gaussian Processes
- Decision Trees
- Ensemble methods (e.g., gradient boosting, random forest, bagging, voting)
- Multiclass and multioutput algorithms

Comparison of classifier algorithms

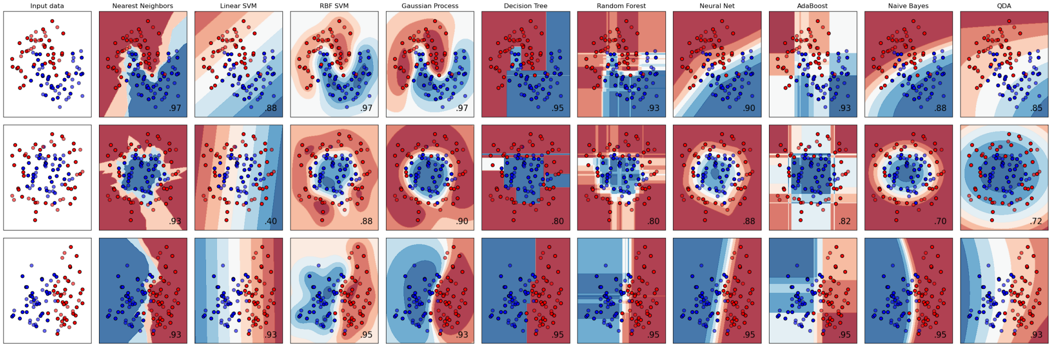

Figure 11: Classifier comparison taken from
https://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html

Supplementary materials

Split dataset and Cross-validation

- CV evaluates how well the model generalize unseen data
- CV aims to prevent model's overfitting
- Common types of CV:
  - k-fold cross validation
  - Leave-One-Out Cross-Validation
  - Stratified k-fold cross validation
- Visual explanation of Cross-Validation

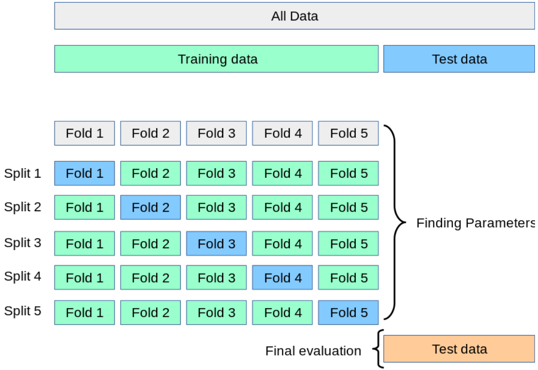

Figure 12: k-fold cross validation illustration. Taken from https://scikit-learn.org

Choosing the right estimator

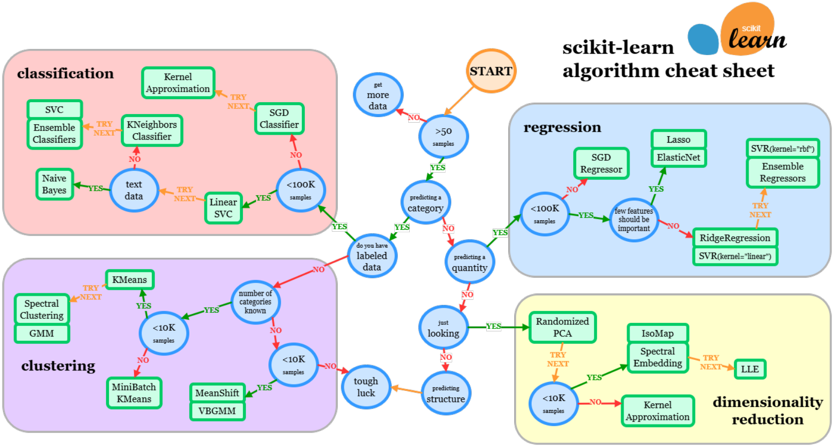

Figure 13: Map of ML taken from https://scikit-learn.org/stable/machine_learning_map.html

Be aware of ML drawbacks

- Data dependency
- Lack of interpretability
- Risk of underfitting and overfitting
- High computational cost
- Security and privacy breach
  - Data poisoning
  - Adversarial attacks
  - Leakage of sensitive information

Figure 14: Taken from https://techgenies.com

Studying ML further?

- Scikit-learn course
- GitHub - microsoft/ML-For-Beginners: 12 weeks, 26 lessons
- Gurobi Machine Learning Manual

Concluding remarks

Concluding remarks

Summary

- Key concepts of ML
- Data preparation techniques
- ML algorithms
  - Unsupervised learning
  - Supervised learning
- Model training and evaluation

Concluding remarks

Summary

- Key concepts of ML
- Data preparation techniques
- ML algorithms
  - Unsupervised learning
  - Supervised learning
- Model training and evaluation

Every dataset tells different story. ML is not just about algorithms but it's more about revealing a story and translating them into impact.

References

References

Cielen, Davy, and Arno Meysman. 2016. *Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools*. Simon and Schuster.

Wang, Wenmin. 2025. *Principles of Machine Learning: The Three Perspectives*. Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-97-5333-8.