Coputational Social Science Methods in R
Introduction to Machine Learning

Taehee Kim

Summer Semester 2022

University of Oldenburg

What is Machine Learning?

Acknowledgement

Recommended Reading

- Raschka, Sebastian, and Vahid Mirjalili. Python machine learning. Packt Publishing Ltd, 2017.

What is Machine Learning?

- Field of study that gives computers the ability to learn without being explicitly programmed (Arthur Samuel, 1959)
- it learns a function that maps an input to an output based on a training dataset (set of examples of input and output)
- Machine learning is not a single approach but rather a diverse array of techniques
- Machine learning techniques include classification, regression, clustering, Bayesian networks, etc...

Examples

- spam filter, image detection, self-driving car, AlphaGo..etc

Definition of Machine Laearning

A machine learning algorithm is an algorithm that can learn from data:
" A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at the task in T, as measured by P, improves with experience E." (Mitchell, T. 1997)

Three elements

- Experience E $\rightarrow$ train the algorithm (model) by maximizing the performance P on the training set E.
- Task T $\rightarrow$ task is solved by the model trained by E.
- Performance P $\rightarrow$ it should increase with E.

Goal of ML

- learning from data
- execute task T based on experience E with optimal performance P

Three types of machine learning

Supervised Learning

- Right answers are given: labeled data
- Direct feedback
- Predict outcome/future
- Two kinds in terms of outputs:

  Regression: predict continuous valued output

  Classification: Predict discrete valued output
- Example: linear/logistic regression, SVM, Neural Network etc..

Unsupervised Learning

- We can't give right answer to all data as the data increase exponentially with the development of technology
- No labeled data
- No feedback
- Find hidden structure in data
- Example: K-means, Gaussian mixture models, PCA etc..

Reinforcement Learning

- Decision process
- Reward system
- Learn series of actions
- Example: Chess game

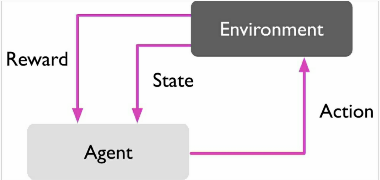


Figure 1: Raschka and Mirjalili (2017)

Supervised vs. Unsupervised

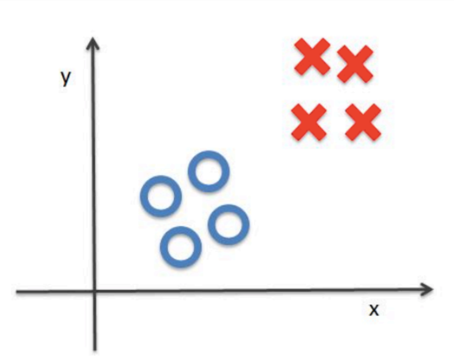
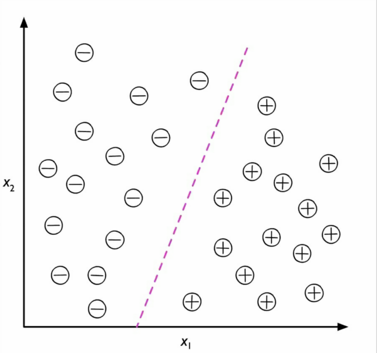


Figure 2: Classification (supervised) vs. Clustering (unsupervised)

Classification vs. Regression



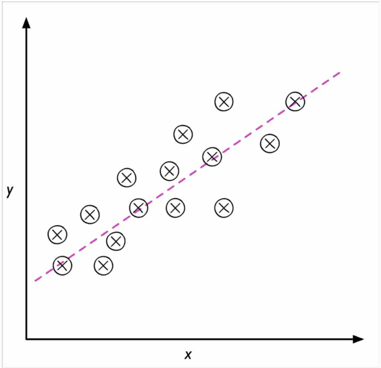


Figure 3: Raschka and Mirjalili (2017)

Process of Supervised Learning

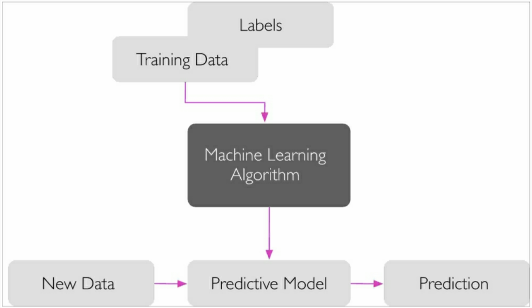


Figure 4: source: Raschka and Mirjalili (2017)

Process of Supervised Learning

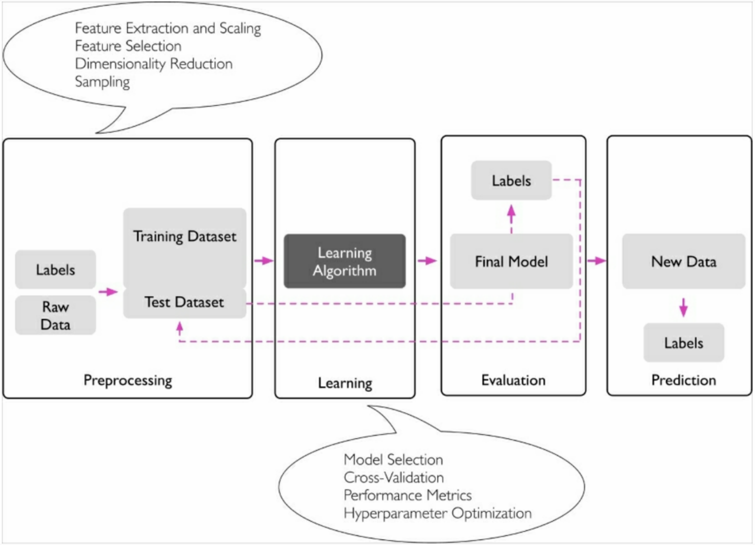


Figure 4: source: Raschka and Mirjalili (2017)

Example: Linear regression with one variable

- Hypothesis: $h_\theta = \theta_0 + \theta_1 x \rightarrow$ model's predict
- Parameters: θ_0, θ_1 $\rightarrow$ we want to find out
- Cost Function: $J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^i) - y^i)^2$
  $\rightarrow$ calculating error, i.e., differences between model's predict and true value
- Goal: $minJ(\theta_0, \theta_1)$ $\rightarrow$ we want to find the parameter which minimize error

How to find $minJ(\theta_0, \theta_1)$?

- Gradient descent:

  Given a function $f(x)$, our objective is: $min_x f(x)$

  Repeat until convergence: $x := x - \alpha\frac{\partial f}{\partial x}$, where α: learning rate

  α is too small, more iterations; α is too large, may not converge
  $\rightarrow$ α is hyperparameter

Gradient descent

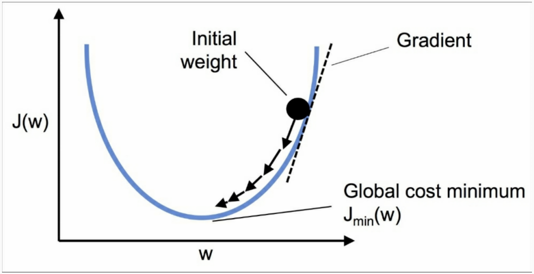


Figure 5: source: Raschka and Mirjalili (2017)

Gradient descent vs. normal equation

- Normal equation:
  – No iterations, but need to compute $(X^TX)^{-1}(X^TY)$. It is slow when n is very large
- Gradient descent:
  – Need many iterations but works well even when number of features n is very large

Hyperparameter

- is a parameter whose value is set before the learning process begins.
- the values of other parameters are derived via training

Grid Search

- The traditional way of performing hyperparameter optimization.
- an exhaustive searching through a manually specified subset of the hyperparameter space of a learning algorithm.

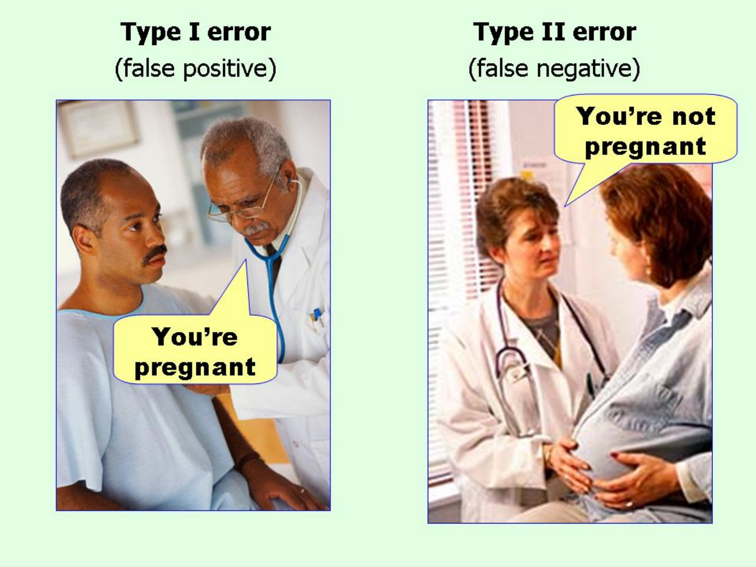


Figure 6: Source of Image: Effect Size FAQs by Paul Ellis

Confusion matrix

		Predicted value	
		Positive	Negative
True value	True	True Positive	False Negative
	False	False Positive	True Negative

Table 1: Confusion matrix: a table visualizing the performance of a supervised learning.

- Accuracy $= \frac{TP+TN}{TP+TN+FN+FP}$
- Precision $= \frac{TP}{TP+FP}$: fraction of true positives among positives
- Recall $= \frac{TP}{TP+FN}$: fraction of true positives among true cases
- You have to be careful especially when your dataset is unbalanced.

Confusion Matrix

How do you assess this classification performance? If this a result of some disease examination, do you want to take this test?

		Predicted value	
		positive	negative
True value	true	61	492
	false	16	1593

- N of dataset = 2062
- N of false = 1509; N of true = 553
- Accuracy: 0.765

Confusion Matrix

How do you assess this classification performance? If this a result of some disease examination, do you want to take this test?

		Predicted value	
		positive	negative
True value	true	61	492
	false	16	1593

- N of dataset = 2062
- N of false = 1509; N of true = 553
- Accuracy: 0.765
- Precision: 0.79; Recall: 0.11

Overfitting

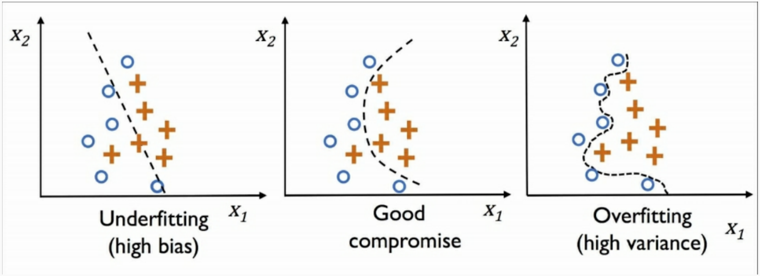


Figure 7: source: Raschka and Mirjalili (2017)

When overfitting occur?

- flexible model with too many parameters
- not enough training data