

Basic Machine Learning Models

Lesson 3

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON



Some Machine Learning References

- General
 - Jiawei Han, [Data Mining: Concepts and Techniques](#), (The Morgan Kaufmann Series in Data Management Systems)
 - Tom Mitchell, *Machine Learning*, McGraw Hill, 1997
 - Christopher Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995
- Adaboost
 - Friedman, Hastie, and Tibshirani, “Additive logistic regression: a statistical view of boosting”, *Annals of Statistics*, 2000
- SVMs
 - <http://www.support-vector.net/icml-tutorial.pdf>

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Overview

- Section 1
 - Supervised and Unsupervised Machine Learning
- Section 2
 - How machine learns
 - Common Pitfalls in Machine Learning
- Section 3: Review
 - Linear Regression (review)
 - Logistic regression (review)
- Section 4
 - Naïve Bayesian Classifier (method required to solve homework)

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Supervised vs Unsupervised Machine Learning

Supervised vs. Unsupervised Learning

- **Supervised learning:** You have a set of independent variables X , and a target variable Y . Your machine learning task is to build a map from $X \rightarrow Y$.
 - Classification and regression models both belong to supervised learning
- **Unsupervised learning (clustering)**
 - Class labels of the data are unknown
 - Given a set of data, the task is to establish the existence of classes or clusters in the data

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

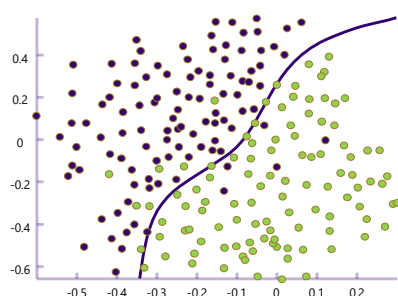
Examples of Supervised Machine Learning

- Fraud transaction: we know which transactions in the training data were fraud (1), which were not (0)
- Readmission: we know which patients were readmitted to hospital within a certain time window after discharge
- Recommendation: we know which items were presented to customers, and which items were clicked, added to cart, or purchased.

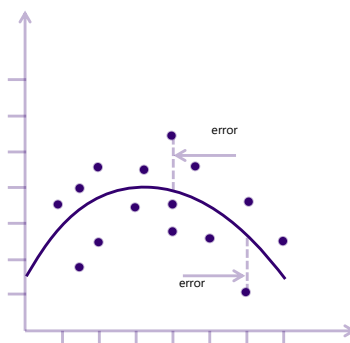
PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Three typical supervised machine learning tasks: Classification, Regression and Recommendation

Classification



Regression



Recommenders



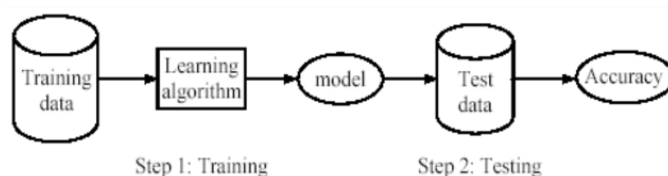
PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Supervised learning process: two steps

Learning (training): Learn a model using the training data

Testing: Test the model using unseen test data to assess the model accuracy

$$Accuracy = \frac{\text{Number of correct classifications}}{\text{Total number of test cases}},$$



PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

What do we mean by learning?

- **Given**

- a data set D ,
- a task T , and
- a performance measure M ,

a computer system is said to **learn** from D to perform the task T if after learning the system's performance on T improves as measured by M .

- In other words, the learned model helps the system to perform T better as **compared to no learning**.

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Fundamental assumption of learning

Assumption: The distribution of training examples is **identical** to the distribution of test examples (including future unseen examples).

- In practice, this assumption is often violated to certain degree.
- Strong violations will clearly result in poor classification accuracy.
- To achieve good accuracy on the test data, training examples must be sufficiently representative of the test data.

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Summary

- >Supervised and Unsupervised Learning
- >Three typical supervised machine learning tasks
- >2 Stages in machine learning
- >Fundamental assumption in machine learning

