Introduction to Machine Learning

Dave Kincaid

3/21/2011

Outline

Types of machine learning

Supervised learning

Learning using examples which have both features and the desired target.

Types of machine learning

Supervised learning

Learning using examples which have both features and the desired target.

Unsupervised learning

Learning using only features. Don't know (or don't provide) the targets

Types of machine learning

Supervised learning

Learning using examples which have both features and the desired target.

Unsupervised learning

Learning using only features. Don't know (or don't provide) the targets

Reinforcement learning

Computer is only given feedback as to whether the answer is right or wrong.

Types of machine learning

Supervised learning

Learning using examples which have both features and the desired target.

Unsupervised learning

Learning using only features. Don't know (or don't provide) the targets

Reinforcement learning

Computer is only given feedback as to whether the answer is right or wrong.

Evolutionary learning

Learning where a solution is evolved from some starting population based on a fitness function.

Problem types

Regression

- The target is a continuous number

Problem types

Regression

- The target is a continuous number

Classification

- Target is a discrete set of classes

Short List of Algorithms

Supervised learning algorithms

- **Naive Bayes**
- **Support Vector Machines (SVM)**
- **k-Nearest Neighbors**
- Decision trees (C4.5)
- Random forests
- Logistic regression
- Stochastic Gradient Descent
- Artificial Neural networks

Unsupervised learning algorithms

- k-means clustering
- Artificial neural networks
- Self-organizing maps
- Hierarchical clustering
- Mean shift clustering
- Affinity propagation

Languages and libraries

Java

- Apache Mahout
- Weka

Python

- Scikit-learn
- PyBrain
- Natural Language Toolkit (NLTK)
- PyML

C#

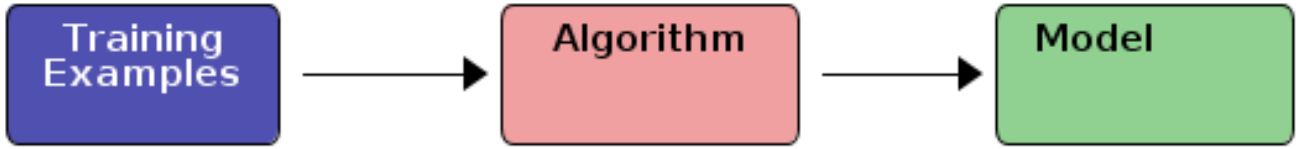
- IKVM & Weka
- AForge.NET & Accord.NET

Others

- R stats package w/various add-ons
- libsvm, libFANN (C/C++)
- Incanter (Clojure)

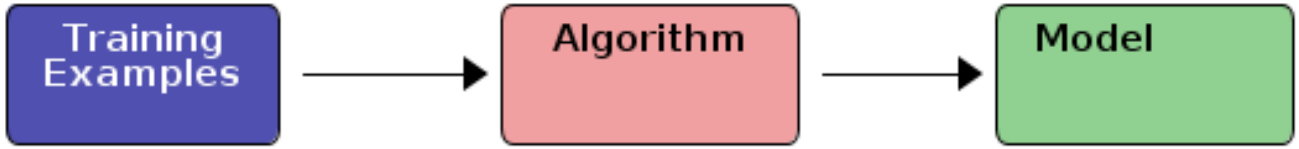
Workflow

- Training the model

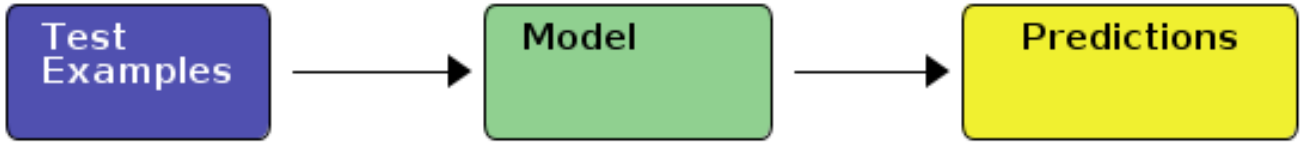


Workflow

- Training the model

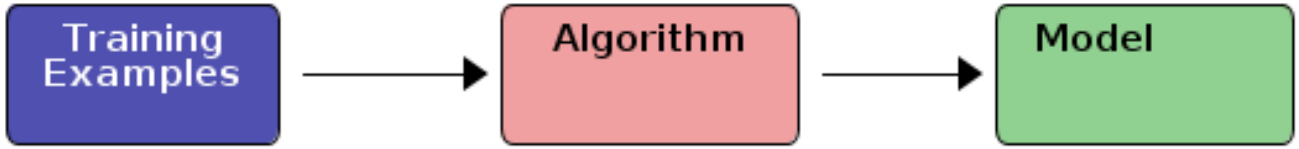


- Testing the model

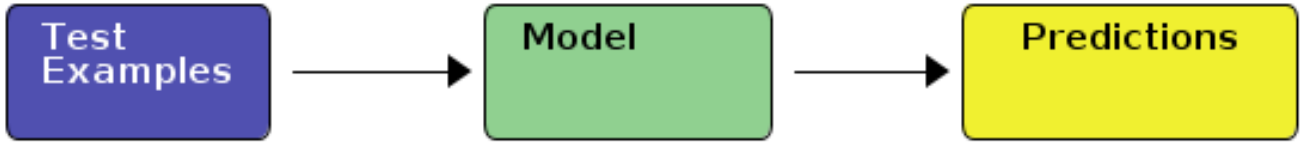


Workflow

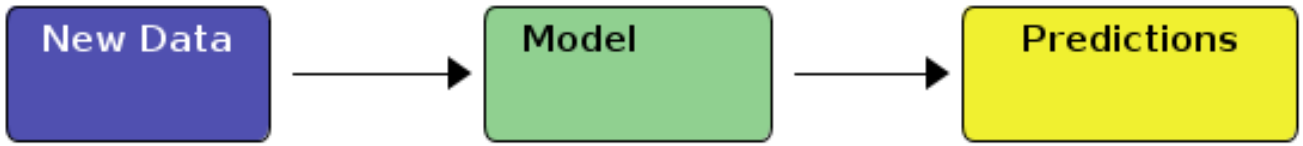
- Training the model



- Testing the model



- Using the model



Species classifier

Example (Species Classifier Example)

- Features: Name, class, sex, age, weight, color, state
- Target: Species

Species classifier

Example (Species Classifier Example)

- Features: Name, class, sex, age, weight, color, state
- Target: Species

Algorithms

- Naive Bayes
- k-Nearest Neighbors
- Support Vector Machine

Species classifier

Example (Species Classifier Example)

- Features: Name, class, sex, age, weight, color, state
- Target: Species

Algorithms

- Naive Bayes
- k-Nearest Neighbors
- Support Vector Machine

Code used

- Python with the Scikit-Learn library

Species Classifier: Load the data

```
import csv

inputfile = open("species.csv")

for i in range(5):
    print i
```

Algorithms: Naive Bayes

High level description of the Naive Bayes algorithm

Species Classifier: Naive Bayes: Train the model

Species Classifier: Naive Bayes: Test the model

Species Classifier: Naive Bayes: Measure the accuracy

Algorithms: k-Nearest Neighbors

High level description of the k-Nearest Neighbors algorithm

Species Classifier: kNN: Train the model

Species Classifier: KNN: Test the model

Species Classifier: kNN: Measure the accuracy

Algorithms: Support Vector Machine

High level description of the Support Vector Machine algorithm

Species Classifier: SVM: Train the model

Species Classifier: SVM: Test the model

Species Classifier: SVM: Measure the accuracy

Testing sandbox