Introduction to Machine Learning

...

What is Machine Learning?

What is Machine Learning?

What is Learning?

- How are we learning in this classroom?

What is Machine Learning?

Tom Mitchell

A computer program (L) is said to learn

- from experience E
- with respect to some class of tasks T
- and performance measure P
- if its performance at tasks in T, as measured by P, improves with experience E.

What is Machine Learning?

Another perspective

- Teaching a computer program how to add numbers
- Teaching a computer program how to cook rice
- Teaching a computer program how to recognise a face or drive a car

Types of Machine Learning

Supervised

- the data comes with a target attribute that we want to predict
- there are two types of Supervised Machine Learning problems
  - Classification: the target variable is categorical
  - Regression: the target variable is numeric

Unsupervised

- The training data consists of a set of input vectors without any corresponding target values
- The goal is to explore the structure of data
- Can be very useful in Supervised Learning

Unsupervised

- **Clustering**: discover groups of similar examples within the data
- **Density Estimation**: determine the distribution of data within the input space
- **Dimension Reduction**: for example, project the data from a high-dimensional space down to two or three dimensions for the purpose of visualization
- **Manifold learning** (nonlinear dimension reduction): Finding a low-dimensional manifold containing the data points

Semi-supervised

- A class of machine learning techniques that make use of both labeled and unlabeled data for training
- Typically a small amount of labeled data with a large amount of unlabeled data.

Semi-supervised

- Labeled data is often costly to generate, whereas unlabeled data is generally not
- Example: collecting images of human faces and then tagging tem with their emotion

Reinforcement Learning

How software agents ought to take actions in an environment so as to maximize some notion of cumulative reward...

Some Terminology

Task (T)

- Classification
- Regression
- Clustering

Learner (L)

- Naive Bayes
- Linear Regression
- Support Vector Machines
- Logistic Regression
- Random Forest
- Gradient Boosting Machines

Performance Measure (P)

- Binary Classification
  - Accuracy
  - Precision
  - Kappa
- Multi-class Classification
  - m-logloss
- Regression
  - RMSE

Experience

(from the Training set)

Training the model with right parameters

Feature Space

- A feature vector is an n-dimensional vector of numerical features that represent some object
- The vector space associated with these vectors is called the feature space
- The space of all possible instances

Feature Space

(examples)

- When representing images, the feature values might correspond to the pixels of an image
- When representing texts, perhaps term occurrence frequencies

Hyperparameter Space

- Parameters vs Hyperparameters
- Hyperparameter determine how the algorithm learns
- Choosing the right values can fix bias
- The vector space associated with all the possible values the hyperparameters can take