

Contents

1	Machine Learning Basics	1
1.1	4 Broad Machine Learning Tasks	1
1.2	Classification	1
1.3	Regression	1
1.4	Classification vs Regression	2
1.5	Supervised or Unsupervised?	2
1.6	Clustering: Unsupervised Learning	2
1.7	Dimensionality Reduction	2
1.8	The Big Picture	3

1 Machine Learning Basics

1.1 4 Broad Machine Learning Tasks

1. Classification
2. Regression
3. Clustering
4. Dimensionality Reduction

1.2 Classification

Broadly, classification is the task of making a decision on "what" something is. (Making a choice between n choices)

There are many examples:

1. Spam Recognition (Spam, or not spam: Binary classification)
2. Face Recognition (This is Matt Damon, etc.: Multi-class classification)

1.3 Regression

Obtain a number from data. This is the task of getting a quantitative number from data.

Examples:

1. Housing price from feature data

1.4 Classification vs Regression

- Regression: labels are continuous and ordered
- Classification: labels are categorical

1.5 Supervised or Unsupervised?

Supervised learning is when you train your model with data that has a "correct answer" or label. ie: Given an image of a face, you also have the name of the owner of the face.

Unsupervised learning is when you train your model with data that has no label. Unsupervised learning is great for training a model that seeks to find common patterns or traits among data points.

- Classification: Supervised
- Regression: Supervised
- Clustering: Unsupervised
- Dimensionality Reduction: Supervised/Unsupervised

1.6 Clustering: Unsupervised Learning

A type of task that seeks to "split" data points by how "similar" they are by their proximity to each other.

- **Input:**

1. large number of data points (n)
2. number of clusters that is much smaller than n ($k \ll n$)

- **Output:**

1. Data points with associated labels found through clustering, where the labels are 1 of the k clusters that was inputted into the training

1.7 Dimensionality Reduction

A task that seeks to eliminate unnecessary data from an input that makes the dimensions smaller and thus easier to be used for computation

1.8 The Big Picture

1. Tasks are the general problems that one seeks to solve.
2. Methods are the approach that we attempt to solve problems (or tasks).
3. Algorithms are the specific, mathematical algorithms that determine results. These algorithms are used to obtain intermediary output that we can tune and adjust to converge on the best possible answer for this specific algorithm/model combination.

Example:

- Regression
- Classification
 - Methods
 - * Neural Networks
 - * SVM
 - * Logistic Regression
 - Algorithms
 - Gradient Descent
 - Stochastic GD
 - Coordinate Descent (CD)
 - Dual CD
 - Newton's Method
 - * Decision Tree
 - * Nearest Neighbor
- Clustering
- Dimensionality Reduction