2. Objectives

Bookmark this page

Introduction to Machine Learning

At the end of this lecture, you will be able to

- understand the goal of machine learning from a movie recommender example
- understand elements of **supervised learning**, and the difference between the **training set** and the **test set**
- understand the difference of **classification** and **regression** - two representative kinds of supervised learning

Feature Vector Demystified 1

1/1 point (graded)

We have a movie recommending system that reads description of each movie and determines some important characteristics of the movie. In particular, it examines whether each of the criterion below is true for that movie:

1. Is it a comedy movie?

2. Is it an action movie?

3. Was the movie directed by Spielberg?

4. Do dinosaurs appear in the movie?

5. Is it a Disney film?

For example, when the recommending system reads descriptions of "Jurassic Park", the answers for the five questions above will be "no, yes, yes, yes, no." On the other hand if the recommending system reads descriptions of "High School Musical", the answers will be "no, no, no, no, yes"

The system converts "yes" into 1, "no" into 0, and makes a feature vector X for each movie. So $X_{JurrasicPark}$ will be $[0,1,1,1,0]$, while $X_{HighSchoolMusical}$ will be $[0,0,0,0,1]$

Question 1: Now we have a comedy movie that is not an action movie, that was not directed by Spielberg, that does not have dinosaurs in it, but was produced by Disney. What is this movie's feature vector?

- (●) $[1,0,0,0,1]$

- ◯ $[0, 0, 1, 0, 0]$
- ◯ $[1, 0, 0, 1, 0]$
- ◯ $[1, 1, 0, 0, 0]$

✔

Solution:

The elements of the feature vector should be $1, 0, 0, 0, 1$, because the answers to the five questions are "yes, no, no, no, yes."

Show answer

Submit You have used 1 of 3 attempts

ℹ Answers are displayed within the problem

Feature Vector Demystified 2

1/1 point (graded)

Question 2: What is the dimension of the feature vector of this movie?

Each feature vector has length 5, so its dimension is 5.

Show answer

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

Training Set vs Test Set 1

1/1 point (graded)

The ultimate goal of our recommending system is to predict whether John will like this movie. Now suppose our movie recommending system knows whether John likes or dislikes the following movies:

	comedy	action	Spielberg	Dinosaur Appearance	Disney	**Liked by John?**
movie 1	0	1	0	0	1	1
movie 2	1	1	1	0	0	-1
movie 3	0	1	0	1	1	1
movie 4	1	1	0	1	0	1

(Like is denoted as 1 and dislike as -1 in the above table) On the other hand, the movie recommender does not know whether John likes the following movies when building the model, but will know them after the model is built:

	comedy	action	Spielberg	Dinosaur Appearance	Disney	**Liked by John?**
movie 5	1	0	0	0	0	Don't know yet
movie 6	0	0	0	0	1	Don't know yet
movie 7	0	0	0	1	1	Don't know yet

Assume that, when John evaluates movies, he only does so based on the five criteria.

Question 1: What is the **label** of movie 1, based on the fact that John likes the movie?

- (●) 1
- ( ) -1

✔

Solution:

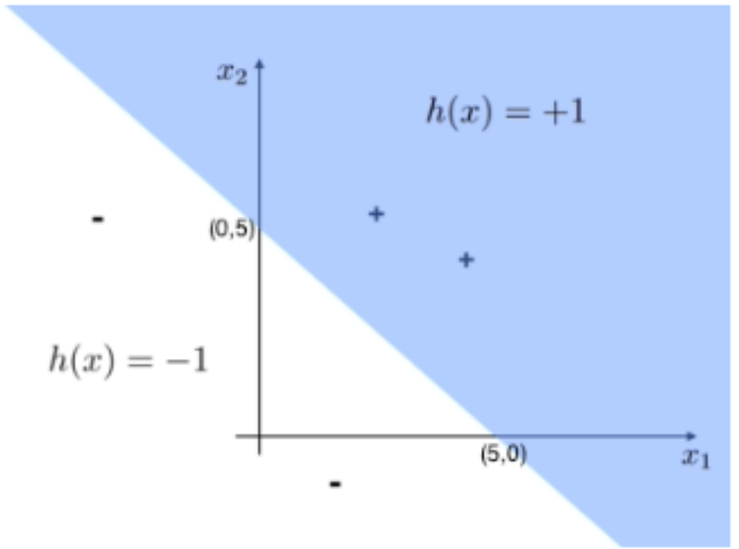
If John likes the movie, the label is 1. Otherwise, it is -1.

Training data can be graphically depicted on a (hyper)plane. **Classifiers** are **mappings** that take **feature vectors as input** and produce **labels as output**. A common kind of classifier is the **linear classifier**, which linearly divides space(the (hyper)plane where training data lies) into two. Given a point x in the space, the classifier h outputs $h(x) = 1$ or $h(x) = -1$, depending on where the point x exists in among the two linearly divided spaces.

Linear Classifier

1/1 point (graded)

We have a linear classifier h that takes in any point on a two-dimensional space. The linear classifier h divides the two-dimensional space into two, such that on one side $h(x) = +1$ and on the other side $h(x) = -1$, as depicted below.



For $x = (10, 10)$, would $h(x)$ be -1 or $+1$?

Suppose a classifier correctly classifies 5 points in the training set and 1 points in the test set. Suppose it incorrectly classifies 5 points in the training set and 2 points in the test set. What is the training error? Is it better than chance?

- (●) 0.5, equal to chance
- ( ) 0.46, worse than chance
- ( ) 0.55, better than chance
- ( ) 0.33, worse than chance

✔

Solution:

We only focus on the training points since the question is asking for training error. We correctly classify 50 percent of points, making this classifier equal to chance.

Show answer

Submit You have used 1 of 3 attempts

ℹ Answers are displayed within the problem

Hypothesis Space

1/1 point (graded)

What is the meaning of the "hypothesis space"?

- ○ the set of test points
- ◉ the set of possible classifiers
- ○ the set of training points
- ○ the positive test examples

✔

Solution:

Each classifier represents a possible “hypothesis" about the data; thus, the set of possible classifiers can be seen as the space of possible hypothesis

Show answer

Submit You have used 2 of 3 attempts

Classification maps **feature vectors** to **categories**. The number of categories need not be two - they can be as many as needed. **Regression** maps feature vectors to **real numbers**. There are other kinds of supervised learning as well.

For a more thorough statistical background on classification and regression, please check out the following links. Classification Regression

---

Classification or Regression? 1

1/1 point (graded)

Question 1: We want to come up with a classifier that classifies each news article into one of the following categories: politics, sports, entertainment. Is this a classification problem or a regression problem?

- (●) classification
- ( ) regression

✔

Solution:

Because we would like to predict the **category** an article would belong to, this problem is a classification problem.

Show answer

Classification or Regression? 2

1/1 point (graded)

Question 2: We want to estimate the price of bitcoin after 30 days. Is this a classification problem or a regression problem?

- ○ classification
- ◉ regression

✔

Solution:

Because we would like to predict the **real** number price of bitcoin, this is a regression problem.

Show answer

Submit You have used 1 of 3 attempts

ℹ Answers are displayed within the problem

Different Types of Learning

Choose the type of learning that best corresponds to each of the following statements.

1)Labelled training and test examples

- (●) supervised learning
- ( ) unsupervised learning
- ( ) semi-supervised learning
- ( ) active learning
- ( ) transfer learning
- ( ) reinforcement learning

✔

2)Using knowledge from one task to solve another task

- ( ) supervised learning
- ( ) unsupervised learning

- semi-supervised learning
- active learning
- transfer learning
- reinforcement learning

✔

3)Learning to navigate a robot

- supervised learning
- unsupervised learning
- semi-supervised learning
- active learning
- transfer learning
- reinforcement learning

4)Deciding which examples are needed to learn

- supervised learning
- unsupervised learning
- semi-supervised learning
- active learning
- transfer learning
- reinforcement learning

✔

5)Data with no annotation

- supervised learning
- unsupervised learning
- semi-supervised learning

6)Training and test examples with limited annotation

- ◯ supervised learning
- ◯ unsupervised learning
- ◉ semi-supervised learning
- ◯ active learning
- ◯ transfer learning
- ◯ reinforcement learning

✔

Solution:

Fully labelled training and test examples corresponds to supervised learning. Limited annotation is semi-supervised learning, and no annotation is unsupervised learning. Using knowledge from one task on another task means you're “transferring" information. Learning how to navigate a robot means learning to act and optimize your actions, or reinforcement learning. Deciding which examples are needed to learn is the definition of active learning.