edX

Course > Unit 1 Linear Classifiers and Generalizations (2 weeks) > Lecture 1. Introduction to Machine Learning > 5. A Concrete Example of a Supervised Learning Task
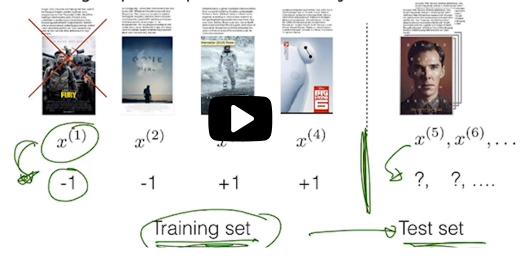
5. A Concrete Example of a Supervised Learning Task
Movie Recommender Problem

That is the task that I wish to solve.

Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

training set

and apply that information to the test set.

So I wish to learn the mapping from examples--

x to labels, plus minus one, on the basis of the training set--

and hope and guarantee that that mapping,

if applied now in the same way to the test examples,

it would work well.

That is the task that I wish to solve.

End of transcript. Skip to the start.

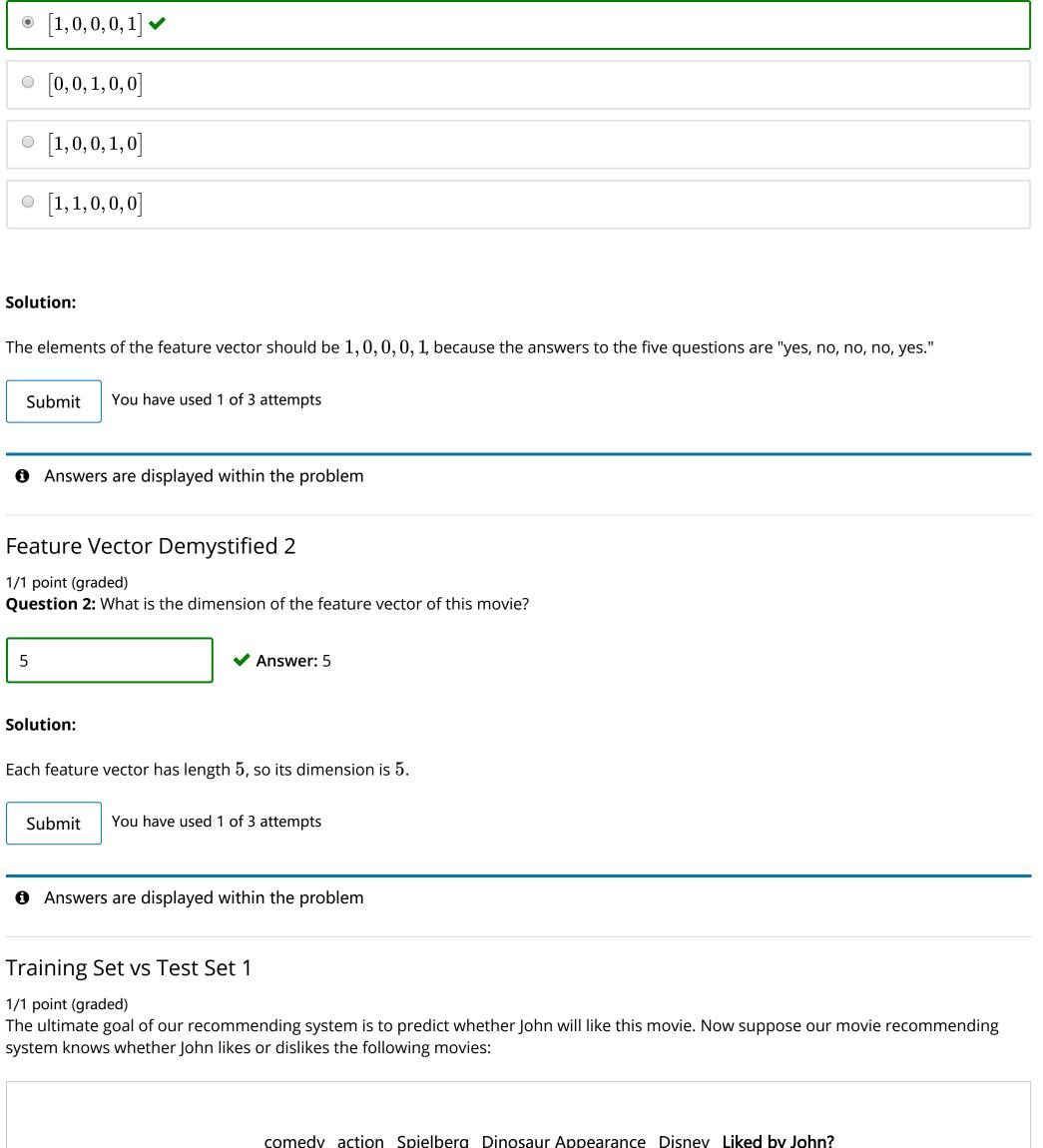## Feature Vector Demystified 1

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."

Submit   You have used 1 of 3 attempts

ⓘ 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?

5   ✔ **Answer:** 5

Solution:

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

Submit   You have used 1 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 has not asked John yet whether he likes the following movies:

	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.

| Submit | You have used 1 of 3 attempts |

ℹ Answers are displayed within the problem

Training Set vs Test Set 2

1/1 point (graded)
Question 2: What movies are in the **training set**? Select all those apply.

☑ movie 1 ✔

☑ movie 2 ✔

☑ movie 3 ✔

☑ movie 4 ✔

☐ movie 5

☐ movie 6

☐ movie 7

✔

Solution:

Movies whose labels are available are in the training set. Thus movies $1, 2, 3, 4$ are in the training set.

| Submit | You have used 1 of 3 attempts |

ℹ Answers are displayed within the problem

Training Set vs Test Set 3

1/1 point (graded)
Question 3: What movies are in the **test set**? Select all those apply.

☐ movie 1

☐ movie 2

- [ ] movie 3
- [ ] movie 4
- [x] movie 5 ✔
- [x] movie 6 ✔
- [x] movie 7 ✔

✔

Solution:

Movies whose labels are not yet available are in the test set. Thus movies $5, 6, 7$ are in the test set. Remember that it is our end goal to predict these movies' labels.

Submit   You have used 1 of 3 attempts

ⓘ Answers are displayed within the problem

Discussion

Topic: Unit 1 Linear Classifiers and Generalizations (2 weeks):Lecture 1. Introduction to Machine Learning / 5. A Concrete Example of a Supervised Learning Task

Show Discussion