

EdX and its Members use cookies and other tracking technologies for performance, analytics, and marketing purposes. By using this website, you accept this use. Learn more about these technologies in the [Privacy Policy](#).



5. A Concrete Example of a Supervised Learning Task Movie Recommender Problem

[Start of transcript. Skip to the end.](#)

A concrete example

▸ Learning to predict preferences from just a little data...



Now, let's take a very concrete example of how to specify a machine learning task and how to solve it.

Let's consider a movie recommender problem. I have a set of movies I've already seen, and I was to use the experience from those movies to make recommendations of whether I would like to see tens of thousands of other movies.

Video
[Download video file](#)

Transcripts
[Download SubRip\(.srt\) file](#)
[Download Text \(.txt\) file](#)

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 $X_{JurassicPark}$ 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 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.

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

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

☒ movie 5 ✓

☒ movie 6 ✓

☒ 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

Hide 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

Add a Post

Show all posts ▼		by recent activity ▼	
	Feature Vector Demystified 2: Helpful links Very helpful for those who are still learning linear algebra: [Linear Algebra - Dimension of a Vector Space][1],[Linear Algebra Toolkit][2],[1]: https://www.youtube.com/watch?v...		4
	That's not what the test set is... I'm sorry, but that's just not the correct usage. The *labeled* training data is split up into several parts. One of those splits is called the test set. I assume we'll cover this, but h...  Community_TA		20
	Is [Staff] planning to post youtube link of the videos ? The videos are not playing smoothly. Can you please post these on youtube.		2
	Three attempts for Training Set vs Test Set 1 Question 1 Seriously? Who will possible need three attempts for this question? ;).		17
	[Typo ?] future movies (not feature movies) 3:27 Which should we users do ? 1) reporting all the suspicious subtitles (some of them can be just misunderstanding of the materials) 2) staying silent about tiny potential typos		3
	Feature Vector Demystified 2/ the wording of the question I'm concerned with the wording of this question. Indeed, when dealing with matrices (as objects), the dimension refers to a couple of integers (number of rows, number of col...		4

Learn About Verified Certificates