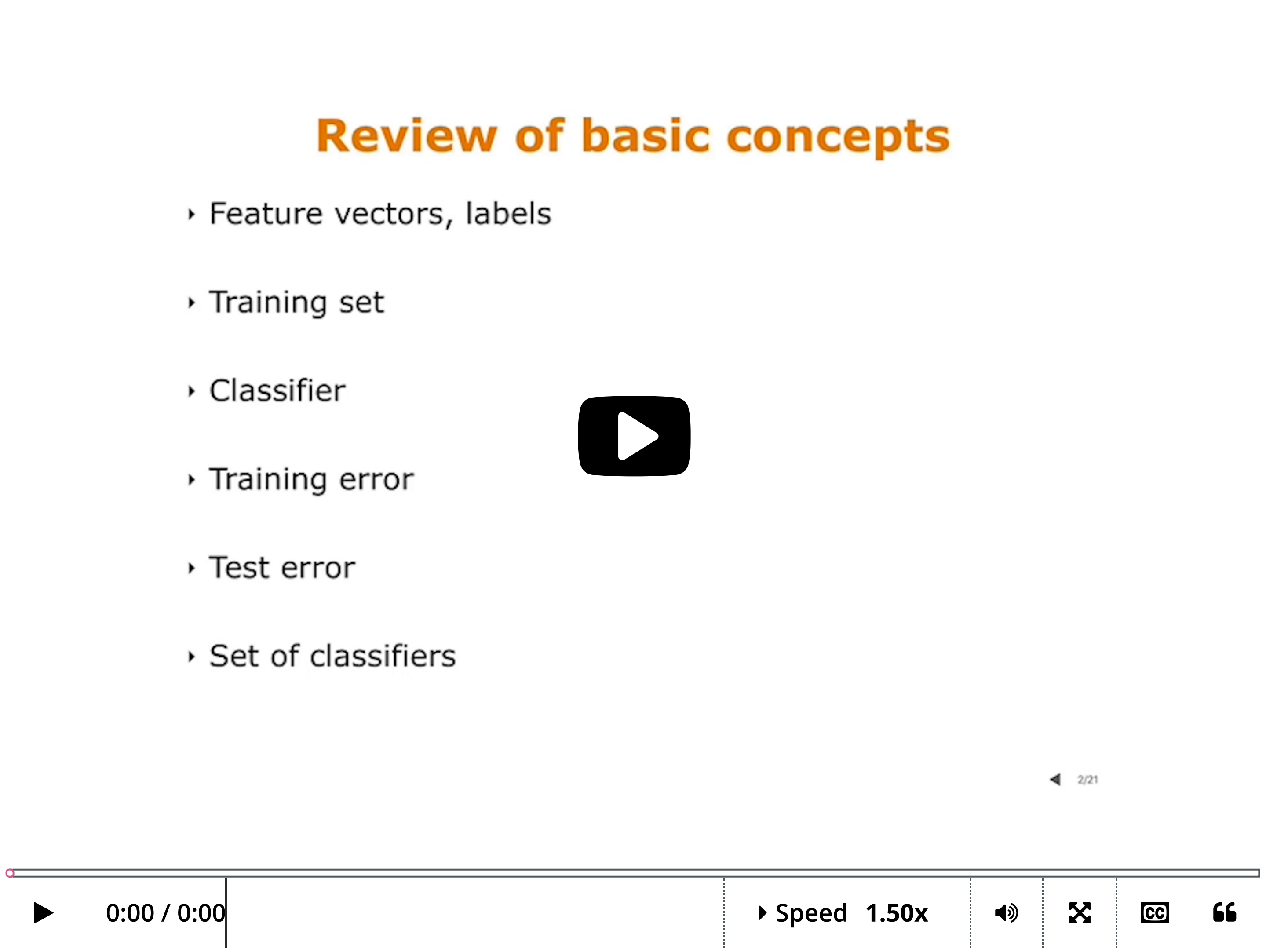
MITx: 6.86x
Machine Learning with Python-From Linear Models to Deep Learning

Help sandipan_dey

Course › Unit 1 Linear Classifiers and Generalizations (2 weeks) › Lecture 2. Linear Classifier and Perceptron › 2. Review of Basic Concepts

2. Review of Basic Concepts

Review of Basic Concepts



Review of basic concepts

- Feature vectors, labels
- Training set
- Classifier
- Training error
- Test error
- Set of classifiers

Start of transcript. Skip to the end.

Welcome back.

This is Machine Learning Lecture Number 2.

Let's start by reviewing basic concepts we've already

seen from lecture number 1.

Feature vectors x provide the context for classifier

to make predictions.

They are vectors.

They belong to Rd So they are d dimensional vectors in general,

Video

Download video file

Transcripts

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

$[[A]]$ either takes value 1 or 0 depending on whether A is True or False. For example, $[[1=3]]=0$, $[[1=1]]=1$, and $[[1\neq 3]]=1$

Concept Review Problem: car accident prediction 1

1/1 point (graded)

In this problem, we will put ourselves in the shoes of a car insurance company. Our goal is to find out whether customers were involved in an accident on July 4th, 1998.

For 8 customers, we know the following information:

1. number of accidents the customer made in the past.
2. number of miles the customer has driven.
3. the customer's age

Also, for 5 of the customers, we know whether each of them was involved in an accident on July 4th, 1998.

If we want to learn a model in a supervised way, what is n, the number of training examples?

$n =$ 5 ✓ **Answer:** 5

Solution:

We have 5 data points with known labels.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

Concept Review Problem: car accident prediction 2

1/1 point (graded)

The insurance company recorded relevant information for all 8 customers, as illustrated in the table below.

number of past accidents	miles customer drove so far	customer's age

customer 1	0	2710.9	21
customer 2	2	13209.2	40
customer 3	1	89001.4	32
customer 4	3	12381.1	18
customer 5	0	1893.5	24
customer 6	2	32493.5	24
customer 7	1	5443.5	30
customer 8	0	4493.5	28

What is the dimension of each feature vector?

$d =$ 3 ✓ **Answer:** 3

Solution:

Each feature vector has length 3 (columns in the table), and thus its dimension is 3.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Concept Review Problem: car accident prediction 3

1/1 point (graded)

How many feature vectors are there in the above table?

$\text{Number of Feature vectors}$ 8 ✓ **Answer:** 8

Solution:

There are 8 rows in the table.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Concept Review Problem: Classifier and Training Error 1

1/1 point (graded)

Assume we have training data and a classifier like the following: (where $h\left(x\right)$ denotes the value outputted by the classifier with the data point as input)

	$h\left(x\right)$	y
data 1	1	1
data 2	-1	1
data 3	1	1
data 4	1	-1
data 5	-1	-1

What is the training error?

$\varepsilon_n\left(h\right) =$ 2/5 ✓ **Answer:** 0.4

Solution:

We have 5 data points total, two of which $h\left(x\right)$ does not match y (data2 and data4). Thus $\varepsilon_n\left(h\right) = \frac{1}{5}\sum_{i=1}^{5}\left[\left[h\left(x_i\right) \neq y\right]\right] = \frac{2}{5}$

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Concept Review Problem: Classifier and Training Error 2

1/1 point (graded)

Now let's examine the training error $\varepsilon_n\left(h\right)$ in a general sense. $\varepsilon_n\left(h\right)$ is a function of: (choose all those apply)

☑ n, the number of training data ✓

☑ h, the classifier ✓

□ the number of test data

✓

Solution:

By definition, $\varepsilon_n\left(h\right) = \frac{1}{n}\sum_{i=1}^{n}\left[\left[h\left(x^i\right) \neq y^i\right]\right]$. Because x, y(training set) is given, $\varepsilon_n\left(h\right)$ depends on n and h. It does no thave any term related to the test data.

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 2. Linear Classifier and Perceptron / 2. Review of Basic Concepts

Add a Post

Show all posts ▼ by recent activity ▼

Issue again with the meaning of test set
As before, there is confusion about the test set. We *do* have outcomes for the test data, else we would not be able to use it for testing. Test data is *not* future as yet unse... 7
Community TA

A doubt regarding notation .
What does the [[A]] symbol means ? 5

About hypothesis space 3

Concept Review Problem: car accident prediction 2/3
Doesn't such description impose to treat columns as vectors, each witch 8 elements ??? As for me "number of past accidents", etc are comonly called attributes, features. 8

Typo
Should it not be "The insurance company recorded relevant information for all 8 customers"? 2

Supervised and semi-supervised Learning

The answer to the first question depends on if it's supervised or semi-supervised learning. So, I think it's good practice for this to be specified, so the reference is clear in the q...

2

Typo in second question

The question should read **"The insurance company recorded relevant information for all 8 customers,"** and not 5.

2

Typo in first question

costumers -> customers

2

Learn About Verified Certificates