Audit Access Expires May 11, 2020
You lose all access to this course, including your progress, on May 11, 2020.
Upgrade by Mar 25, 2020 to get unlimited access to the course as long as it exists on the site.
Upgrade now

2. Higher Order Feature Vectors
Higher Order Feature Vectors

▶ 0:00 / 0:00 | ▶ 1.50x   ◄))   ✕   CC   "

Video
Download video file

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

We can use linear classifiers to make non-linear predictions. The easiest way to do this is to first map all the examples $x \in \mathbb{R}^d$ to different feature vectors $\phi(x) \in \mathbb{R}^p$ where typically p is much larger than d. We would then simply use a linear classifier on the new (higher dimensional) feature vectors, pretending that they were the original input vectors. As a result, all the linear classifiers we have learned remain applicable, yet produce non-linear classifiers in the original coordinates.

There are many ways to create such feature vectors. One common way is to use polynomial terms of the original coordinates as the components of the feature vectors. We have seen two examples in the video above. We will recall the 1-dimensional example here and see another 2-dimensional example in the problem below.

Example: Given 3 training examples with $x^{(t)} \in \mathbb{R}$ $(t = 1, 2, 3)$ that are not linearly separable in 1-dimensional space as shown below,
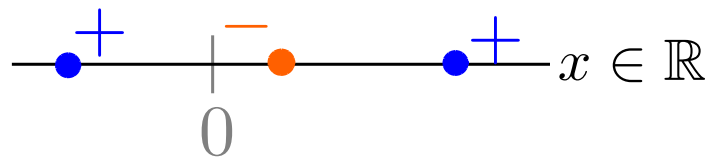
define the feature map $\phi(x) = [\phi_1(x), \phi_2(x)]^T = [x, x^2]^T$, which maps each example $x^{(t)}$ in 1-dimensional space to a corresponding feature vector $\phi\left(x^{(t)}\right) = \left[x^{(t)}, \left(x^{(t)}\right)^2\right]^T$ in 2-dimensional space.

Then, instead of using $\left(x^{(t)}, y^{(t)}\right)$, we use $\left(\phi\left(x^{(t)}\right), y^{(t)}\right)$ as the training examples (where $y^{(t)}$ are the labels):
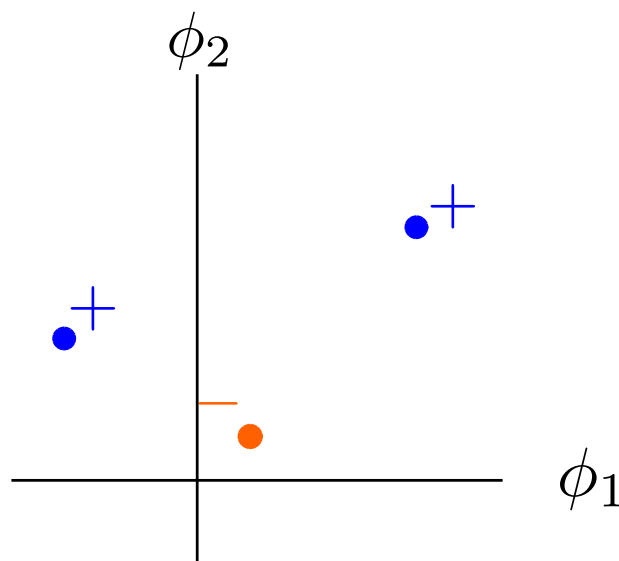
The new training set is linearly separable in the 2-dimensional (ϕ_1, ϕ_2)-space, and we can train a "linear classifier" that is linear in the ϕ-coordinates.

Another 2-Dimensional Example

1/1 point (graded)
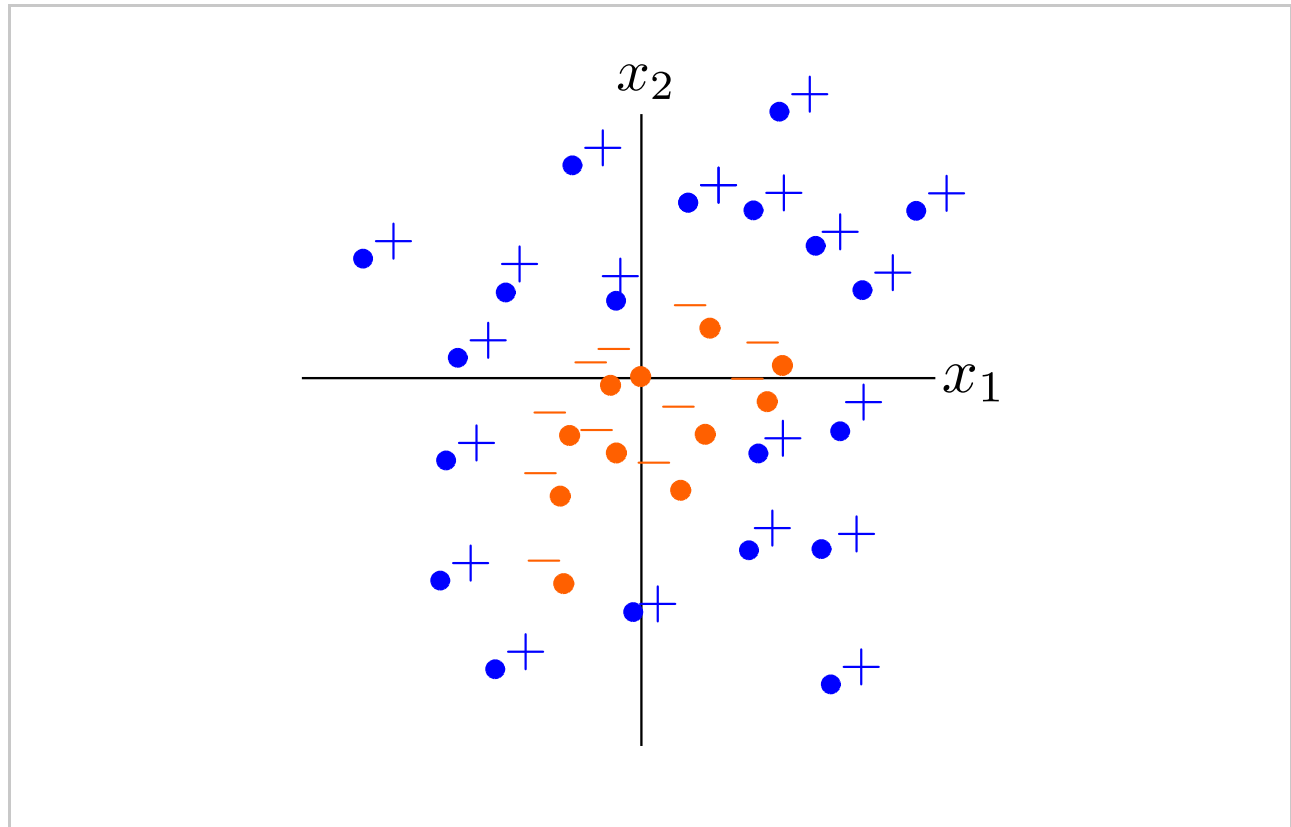
Given the training examples with $x = \left[x_1^{(t)}, x_2^{(t)} \right] \in \mathbb{R}^2$ above, where a boundary between the positively-labeled examples and the negatively-labeled examples is an ellipse , which of the following feature vector(s) $\phi(x)$ will guarantee that the training set $\left\{ \left(\phi(x^{(t)}), y^{(t)} \right), t = 1, \ldots, n \right\}$ (where $y^{(t)}$ are the labels) are linearly separable? (Choose all that apply.)

☐ $\phi\left(x\right)=\left[x_1,x_2\right]^T$

☐ $\phi\left(x\right)=\left[x_1,x_2,x_1x_2\right]^T$

☐ $\phi\left(x\right)=\left[x_1,x_2,x_1^2+x_2^2\right]^T$

☐ $\phi\left(x\right)=\left[x_1,x_2,x_1^2+2x_2^2\right]^T$

☑ $\phi\left(x\right)=\left[x_1,\ x_2,\ x_1x_2,\ x_1^2,\ x_2^2\right]^T$

✔

Solution:

Since a possible boundary is an elipse, and we recall from geometry that the equation of any ellipse can be given as

$$\theta_1x_1+\theta_2x_2+\theta_3x_1x_2+\theta_4x_1^2+\theta_5x_2^2+\theta_0\ =\ \theta\cdot\left[x_1,\ x_2,\ x_1x_2,\ x_1^2,\ x_2^2\right]^T+\theta_0=0,$$

for some $\theta=\left[\theta_1,\theta_2,\theta_3,\theta_4,\theta_5\right]$, we see that defining the feature map to be $\phi\left(x\right)=\left[x_1,x_2,x_1x_2,x_1^2,x_2^2\right]^T$ (the last choice) will allow us to write a linear decision boundary in the ϕ-coordinates:

$$\theta\cdot\phi\left(x\right)+\theta_0\ =\ 0\iff\theta\cdot\left[x_1,x_2,x_1x_2,x_1^2,x_2^2\right]^T+\theta_0\ =\ 0.$$

Let us examine the first three choices in order:

- The first choice $\phi\left(x\right)=\left[x_1,x_2\right]^T$ maps x to x itself in $\mathbb{R}^2$, so the training set remains the same and not linearly-separable.

- The second choice $\phi\left(x\right)=\left[x_1,x_2,x_1^2+x_2^2\right]^T$ gives decision boundaries of the form:

$$\theta\cdot\left[x_1,x_2,x_1^2+x_2^2\right]^T+\theta_0\ =\ \theta_1x_1^2+\theta_2x_2^2+\theta_3\left(x_1^2+x_2^2\right)+\theta_0\ =\ 0,$$

which are circles in the $(x_1,x_2)-$ plane. Hence, if the decision boundary in the $(x_1,x_2)-$ plane is an elipse that is not circular, this choice of ϕ will not give a linear decision boundary in the ϕ-coordinates.

- The argument for the third choice is analogous to the second choice above.

Submit   You have used 3 of 3 attempts

ℹ Answers are displayed within the problem

Discussion   [Hide Discussion]

Topic: Unit 2 Nonlinear Classification, Linear regression, Collaborative
Filtering (2 weeks):Lecture 6. Nonlinear Classification / 2. Higher Order
Feature Vectors

Add a Post

Show all posts	by recent activity

💬 **Very steep learning curve**
How to interpret the last transformation with 5 elements. as opposed to the others having max 3 elements. Feel ...
11

💬 **Magic! (animation on how this high order features work)**
Just want to share this awesome short video on how this stuff work, I think it is brilliant! https://youtu.be/a7cofm...
7

💬 **Segment 6.2 Notes**
👤 Community TA
5

💬 **Absolute value vs. squared?**
Just curious if there is any rationale to the use of a squared value compared to just using the absolute value. Bot...
3

💬 **(many) Transcription errors**
0:33 "carries a computational disadvantage" instead of "carries a computational advantage" 2:36: sign function n...
👤 Community TA
2

☑ **How to choose phi() ?**
While for few data points in a 1-2 dimensions one can have a "look" on the data, things become harder on highe...
👤 Community TA
2

☑ **Guidelines in the Video**
3

💬 **Statistical Power Interpretation?**
I noticed the professor uses the word "power" in classifying the data, and I was wondering if there is a relation o...
2

💬 **Eclipse etc**
In the exercise, above, we first get a clue that the shape of the line that seems best to separate the data in the ini...
2

💬 **Seeming contradictions in explanation to solution**
1

💬 **Solution needs an update**

I think the solution to the question requires an update. Without giving away the answer, I would like to say that t...

2

[STAFF] Will you add the slides to the resources section?

Learn About Verified Certificates