edX

Unit 2 Nonlinear Classification,
Linear regression, Collaborative
Course > Filtering (2 weeks) > Lecture 6. Nonlinear Classification > 2. Higher Order Feature Vectors

2. Higher Order Feature Vectors
Higher Order Feature Vectors

(Caption will be displayed when you start playing the video.)
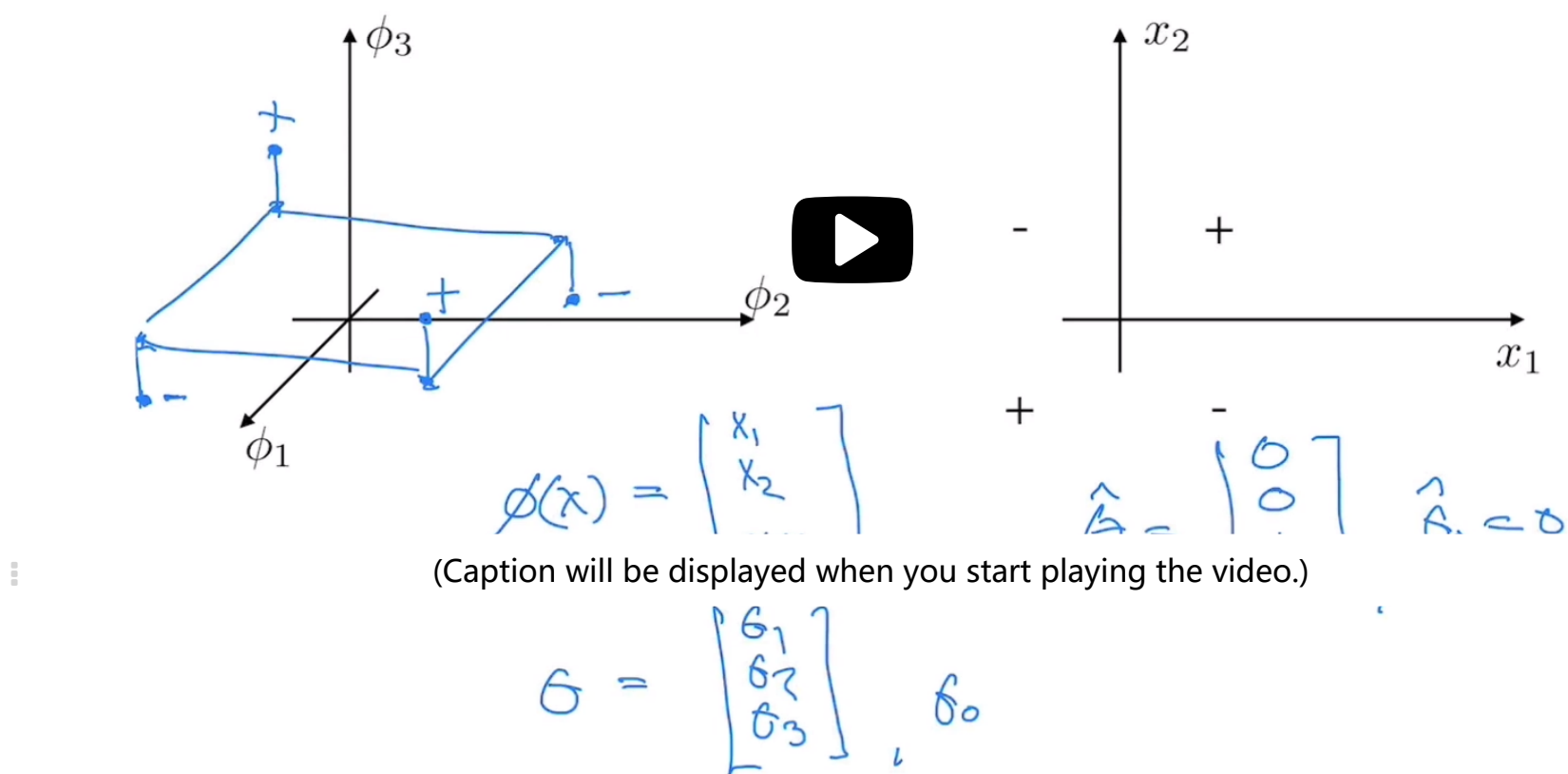
▶   13:57 / 13:57   ▸ 1.0x   ◀)))   ⬚

**Video**
Download video 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,
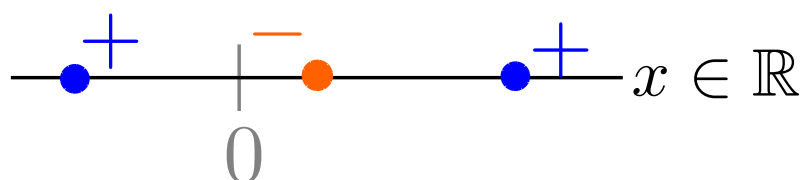
define the feature map $\phi(x) = [\phi_1(x), \phi_2(x)]^T = [x, x^2]^T$, which maps each example $x^{(t)}$ in 1-dimensional space to a corresponding feature vector $\phi(x^{(t)}) = [x^{(t)}, (x^{(t)})^2]^T$ in 2-dimensional space.
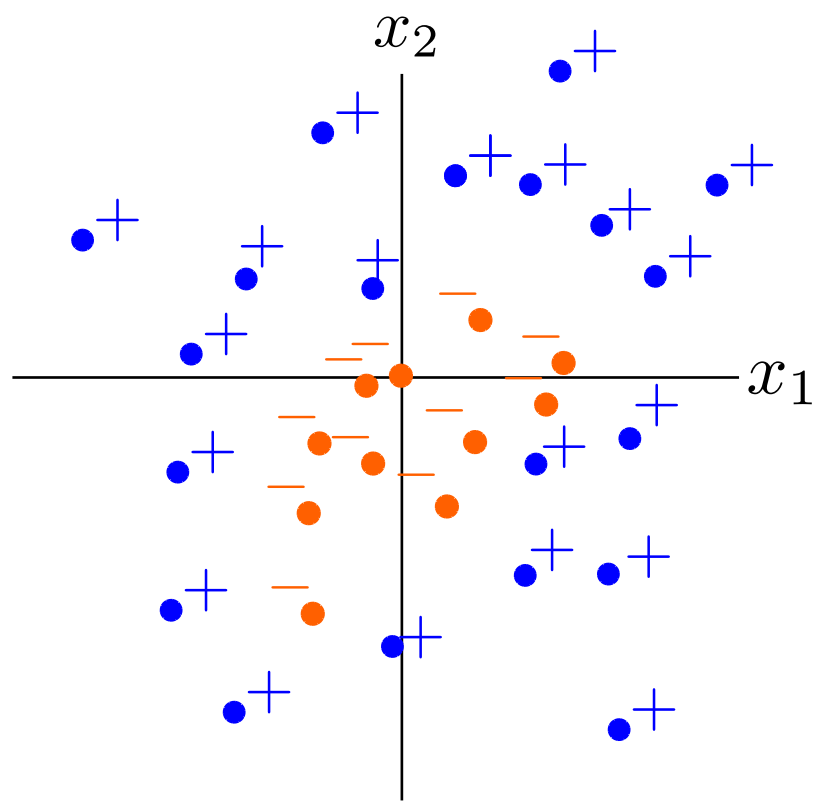
Then, instead of using $\left(x^{(t)}, y^{(t)}\right)$, we use $\left(\phi\left(x^{(t)}\right), y^{(t)}\right)$ as the training examples (where $y^{(t)}$ are the labels):
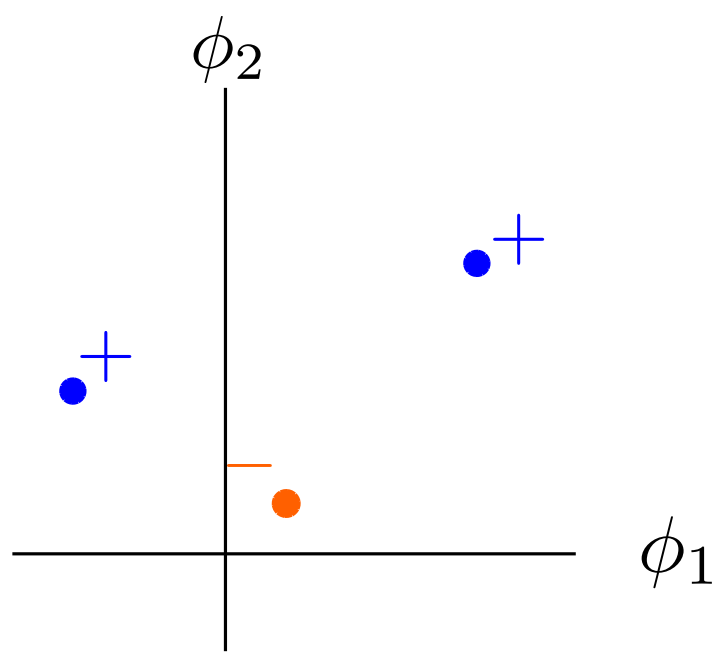
The new training set is linearly separable in the 2-dimensional (ϕ_1, ϕ_2)-space, and we can train a "linear classifier" that is linear in the ϕ-coordinates.

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) $\phi\left(x\right)$ will guarantee that the training set $\left\{\left(\phi\left(x^{(t)}\right), y^{(t)}\right), t = 1, \ldots, n\right\}$ (where $y^{(t)}$ are the labels) are linearly separable?
(Choose all that apply.)

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

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

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

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

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

✔

Solution:

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

$$\theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 + \theta_0 \; = \; \theta \cdot [x_1, \, x_2, \, x_1 x_2, \, x_1^2, \, x_2^2]^T \; = \; 0,$$

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

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

Let us examine the first three choices in order:

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

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

$$\theta \cdot [x_1, x_2, x_1^2 + x_2^2]^T \; = \; \theta_1 x_1^2 + \theta_2 x_2^2 + \theta_3 \left(x_1^2 + x_2^2\right) + \theta_0 \; = \; 0,$$

  which are circles in the (x_1, x_2) − plane. Hence, if the decision boundary in the (x_1, x_2) − <mark>plane is an elipse that is not circular</mark>, 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

ⓘ Answers are displayed within the problem

Discussion

[Show Discussion]

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