✕

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

3. Motivation
Review: True or False

2/2 points (graded)
Consider the classification decision rule

$$\hat{y} = \text{sign}\left(\theta \cdot \phi\left(x\right)\right)$$

where $x \in \mathbb{R}^d$ represent input data and $y \in \{1, -1\}$ is the corresponding predicted labels, and we have omitted the bias/offset term for simplicity.

Given the model above, determine if the following statements are True or False.

1. The feature map ϕ is function from $\mathbb{R}^d$ to $\mathbb{R}^d$.

   | false   ▾ | ✔ **Answer:** false

2. If $\phi\left(x\right) \in \mathbb{R}^D$, then the classification parameter θ is also a vector in $\mathbb{R}^D$. (Answer based on the model as written.)

   | true   ▾ | ✔ **Answer:** true

Solution:

1. The output of feature map $\phi\left(x\right)$ is a vector that is **not** necessarily and often not of the same dimension of the input x. For example, consider $x = \left[\, x_1, x_2 \,\right]^T \in \mathbb{R}^2$, and $\phi\left(x\right)$ a quadratic feature map $\phi\left(x\right) = \left[\begin{array}{ccccc} x_1 & x_2 & x_1^2 & x_1 x_2 & x_2^2 \end{array}\right]^T \in \mathbb{R}^5$.

2. The classification parameter θ must of the same dimension of the feature vectors $\phi\left(x\right)$ for the dot product $\theta \cdot \phi\left(x\right)$ to make sense.

Submit   You have used 1 of 1 attempt
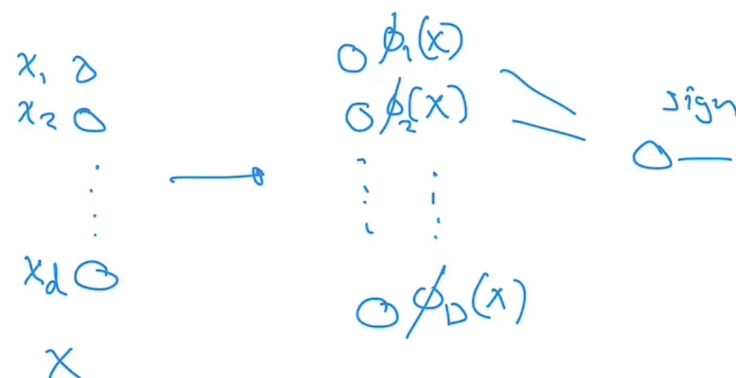
ℹ   Answers are displayed within the problem

Motivation

This could be, for example, that polynomial expansion

as we've been talking about.

Once we have this feature representation,

then we take a linear combination of these coordinate

values to produce the classification decision.

So I'm going to introduce a node here

for these final operation that takes

the sine function of a weighted combination of these features--

and we have weights that are now a theta 1, theta 2, all the way

to theta D. As many coordinates of theta as there

are coordinates in the expanded feature representation.

Now, one particular aspect of this way

of solving a nonlinear classification problem

is that how input x is mapped to a longer feature

representation, 5 of x, that mapping is fixed.

2:45 / 5:01   ▸ Speed 1.50x   🔊 ✕ CC "

Video
Download video file

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

Motivation to Neural Networks

So far, the ways we have performed non-linear classification involve either first mapping x explicitly into some feature vectors $\phi\left(x\right)$, whose coordinates involve non-linear functions of x, or in order to increase computational efficiency, rewriting the decision rule in terms of a chosen kernel, i.e. the dot product of feature vectors, and then using the training data to learn a transformed classification parameter.

However, in both cases, the feature vectors are chosen. They are not learned in order to improve performance of the classification problem at hand.

Neural networks, on the other hand, are models in which the feature representation is learned jointly with the classifier to improve classification performance.

Discussion

<div style="float:right">

Hide Discussion

</div>

Topic: Unit 3 Neural networks (2.5 weeks):Lecture 8. Introduction to Feedforward Neural Networks / 3. Motivation

Add a Post

Show all posts ▼   by recent activity ▼

There are no posts in this topic yet.

✖

Learn About Verified Certificates