3. Motivation
Review: True or False

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

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

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 $x = \begin{bmatrix} x_1, x_2 \end{bmatrix}^T \in \mathbb{R}^2$, and $\phi\left(x\right)$ a quadratic feature map $\phi\left(x\right) = \begin{bmatrix} x_1 & x_2 & x_1^2 & x_1 x_2 & x_2^2 \end{bmatrix}^T \in \mathbb{R}^5$.

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

ⓘ   Answers are displayed within the problem

Motivation

Trying to adjust the feature representation together

with how it is exercised towards classification task.

So this task is quite a bit harder,

and learning these complicated deep architectures will

be difficult. But we will succeed

in doing that even with simple stochastic the gradient

descent method.

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 $\phi(x)$, 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="text-align: right;">Show Discussion</div>

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