<u>Unit 2 Nonlinear Classification,</u> <u>Linear regression, Collaborative</u>

Course > Filtering (2 weeks)

4. Motivation for Kernels:

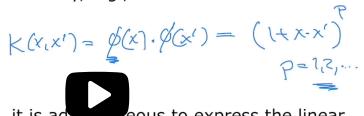
> <u>Lecture 6. Nonlinear Classification</u> > Computational Efficiency

# 4. Motivation for Kernels: Computational Efficiency Motivation for Kernels: Computational Efficiency



# **Kernels vs features**

 For some feature maps, we can evaluate the inner products very efficiently, e.g.,



 In those cases, it is advantageous to express the linear classifiers (regression methods) in terms of kernels rather than explicitly constructing feature vectors

sign (6. p(x)+60) \_0 K(x,x')

that you've already learned.

and tries to classify that-- that's how a linear classifier.

Somehow, we wish to turn that interclassifier

that only depends on those inner products operates in terms of kernels.

And we'll do that in the context of kernel perception

just for simplicity.

But it applies to any linear method

that you've already learned.

**5:34 / 5:34** 

▶ 1.0x

**X** 

End of transcript. Skip to the start.

**Video**Download video file

Transcripts

<u>Download SubRip (.srt) file</u>

<u>Download Text (.txt) file</u>



## Kernels as Dot Products 1

1/1 point (graded)

Let us go through the computation in the video above. Assume we map x and  $x'\in\mathbb{R}^2$  to feature vectors  $\phi\left(x
ight)$  and  $\phi\left(x'
ight)$  given by

$$\phi \left( x 
ight) \ = \left[ {{x_1},\,{x_2},\,{x_1}^2},\,\sqrt 2 {x_1}{x_2},\,{x_2}^2 
ight]$$

$$\phi\left(x'
ight) \ = \left[x_1',\, x_2',\, {x_1'}^{\,2},\, \sqrt{2} x_1' \, x_2',\, {x_2'}^{\,2}
ight].$$

Which of the following equals the dot product  $\phi\left(x\right)\cdot\phi\left(x'\right)$ ?

 $x \cdot x'$ 

$$lacksquare x \cdot x' + (x \cdot x')^2 \checkmark$$

$$(x \cdot x')^2$$

$$2(x \cdot x')^2$$

None of the above

#### Solution:

Expand  $\phi(x) \cdot \phi(x')$  to get

$$egin{array}{lll} \phi\left(x
ight)\cdot\phi\left(x'
ight) &=& x_{1}x'_{1}+x_{2}x'_{2}+x_{1}{x'_{1}}^{2}+2x_{1}x'_{1}x_{2}x'_{2}+x_{2}{x'_{2}}^{2} \ &=& \left(x_{1}x'_{1}+x_{2}x'_{2}
ight)+\left(x_{1}x'_{1}+x_{2}x'_{2}
ight)^{2} \ &=& x\cdot x'+\left(x\cdot x'
ight)^{2}. \end{array}$$

**Remark:** Notice the coefficient  $\sqrt{2}$  of the  $x_1x_2$  terms is necessary for rewriting  $\phi\left(x\right)\cdot\phi\left(x'\right)$  as the function above of  $x\cdot x'$ .

Submit

You have used 1 of 1 attempt

## • Answers are displayed within the problem

#### Kernels as Dot Products 2

1/1 point (graded)

Which of the following feature vectors  $\phi(x)$  produces the kernel

$$K\left(x,x'
ight) \,=\, \phi\left(x
ight)\cdot\phi\left(x'
ight) \,=\, x_{1}x'_{1} + x_{2}x'_{2} + x_{3}x'_{3} + x_{2}x'_{3} + x_{3}x'_{2}$$

(Choose all that apply.)

$$igoplus \phi\left(x
ight) = igl[x_1, x_2, x_3igr]$$

$$\bigcirc \ \phi\left(x
ight) = \left\lceil x_1 + x_2 + x_3 
ight
ceil$$

$$ullet \phi\left(x
ight) = \left[x_1, x_2 + x_3
ight] oldsymbol{\checkmark}$$

$$igoplus \phi\left(x
ight) = igl[x_1+x_3,x_1+x_2igr]$$

#### **Solution:**

Directly expand to see the answer. The fact that there are mixed terms in the kernel, e.g.  $x_2x_3'$ , indicates that some coordinates of the feature vector must be mixed, i.e. contain different  $x_i$ 's.

Submit

You have used 1 of 1 attempt

## • Answers are displayed within the problem

## Discussion

**Show Discussion** 

**Topic:** Unit 2 Nonlinear Classification, Linear regression, Collaborative Filtering (2 weeks):Lecture 6. Nonlinear Classification / 4. Motivation for Kernels: Computational Efficiency