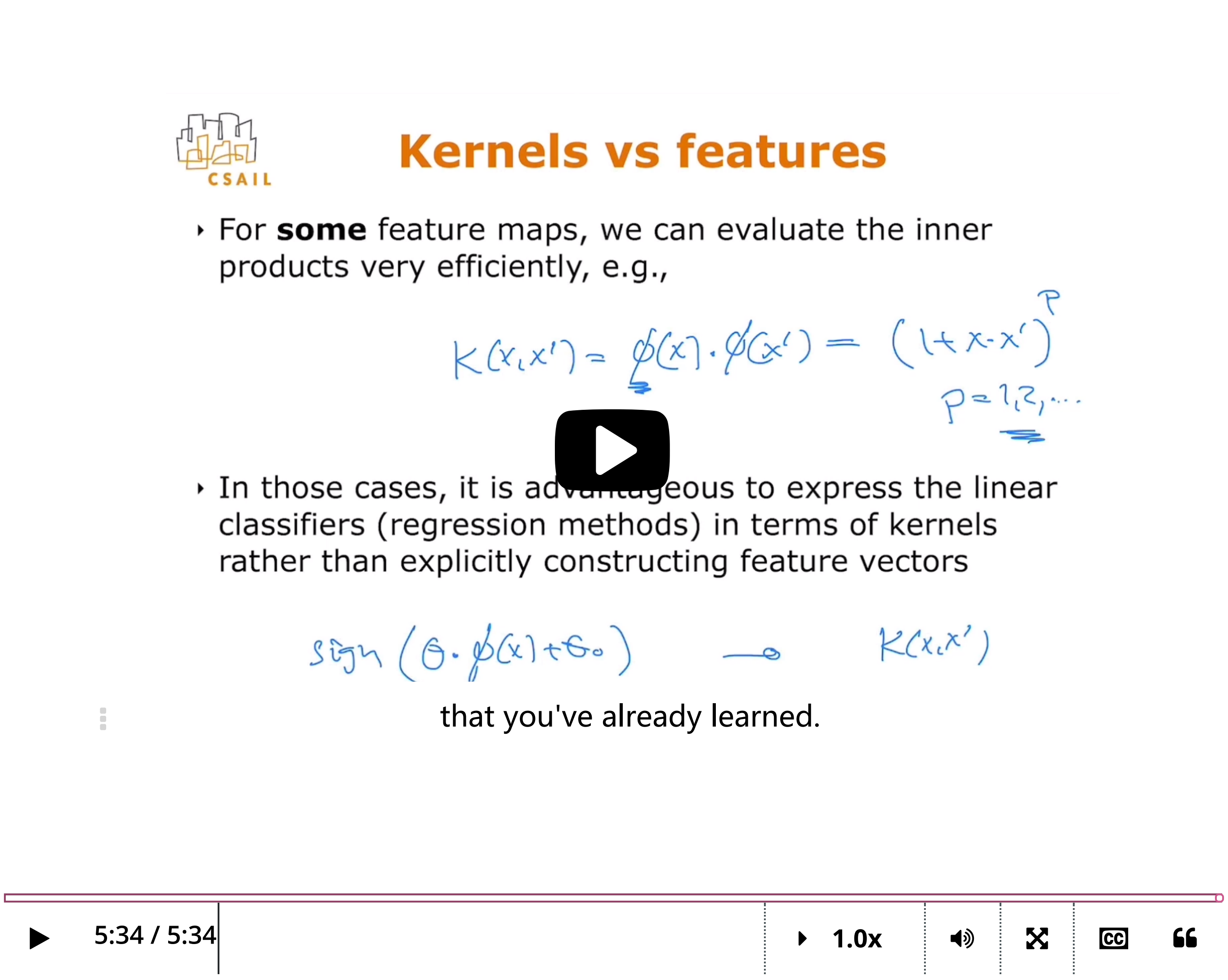
4. Motivation for Kernels: Computational Efficiency

Motivation for Kernels: Computational Efficiency



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

Somehow, we wish to turn that inter- classifier

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.

End of transcript. Skip to the start.

Video
Download video file

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

xuetangX.com 学堂在线

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 $\phi(x)$ and $\phi(x')$ given by

$$\begin{aligned}\phi(x) &= [x_1, x_2, {x_1}^2, \sqrt{2}x_1x_2, {x_2}^2]\\ \phi(x') &= [x'_1, x'_2, {x'_1}^2, \sqrt{2}x'_1x'_2, {x'_2}^2].\end{aligned}$$

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

- ○ $x\cdot x'$
- ◉ $x\cdot x' + (x\cdot x')^2$ ✔
- ○ $(x\cdot x')^2$
- ○ $2(x\cdot x')^2$
- ○ None of the above

Solution:

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

$$\begin{aligned} \phi(x) \cdot \phi(x') &= x_1 x_1' + x_2 x_2' + x_1 {x_1'}^2 + 2 x_1 x_1' x_2 x_2' + x_2 {x_2'}^2 \\ &= (x_1 x_1' + x_2 x_2') + (x_1 x_1' + x_2 x_2')^2 \\ &= x \cdot x' + (x \cdot x')^2. \end{aligned}$$

Remark: Notice the coefficient $\sqrt{2}$ of the $x_1 x_2$ terms is necessary for rewriting $\phi(x) \cdot \phi(x')$ 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

$$K(x, x') = \phi(x) \cdot \phi(x') = x_1 x_1' + x_2 x_2' + x_3 x_3' + x_2 x_3' + x_3 x_2'$$

(Choose all that apply.)

- $\phi(x) = [x_1, x_2, x_3]$
- $\phi(x) = [x_1 + x_2 + x_3]$
- $\phi(x) = [x_1, x_2 + x_3]$ ✔
- $\phi(x) = [x_1 + x_3, x_1 + x_2]$

Solution:

Directly expand to see the answer. The fact that there are mixed terms in the kernel, e.g. $x_2 x_3'$, 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