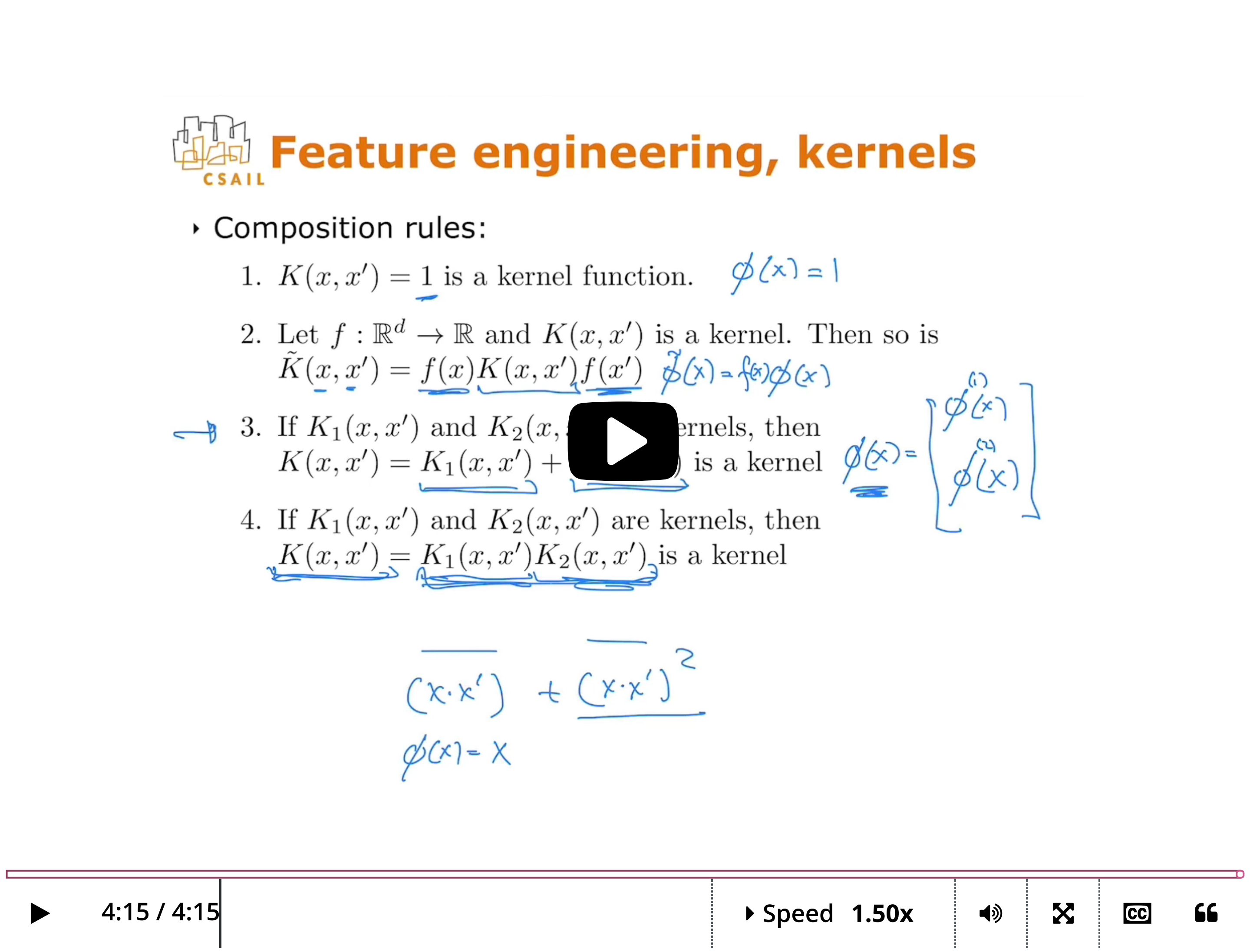
6. Kernel Composition Rules

Kernel Composition Rules



That's called a linear kernel, where phi of x

is just identity.

We can add to it a squared term.

And to verify that this resulting thing has a feature

representation, this squared now is a product of two kernels--

identical kernels.

You get that kernel value squared.

And you add here two kernels together.

And according to the third rule, you get a value kernel

as a result, and so on.

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Kernel Composition Rules 1

1/1 point (graded)

Recall from the video above that if $f:\mathbb{R}^d \to \mathbb{R}$ and $K(x,x')$ is a kernel, so is

$$\widetilde{K}(x,x') = f(x)\,K(x,x')\,f(x')\,.$$

If there exists $\phi(x)$ such that

$$K(x,x') = \phi(x)\cdot\phi(x')\,,$$

then which of the following φ gives

$$\widetilde{K}(x,x') = \varphi(x)\cdot\varphi(x')?$$

- $\varphi(x) = f(x)\,K(x,x)$
- $\varphi(x) = f(x)\,K(x,x')$
- $\varphi(x) = f(x)$
- $\varphi(x) = f(x)\,\phi(x)$ ✔

Solution:

As $f(x), f(x') \in \mathbb{R}$, we have $(f(x)\,\phi(x))\cdot(f(x')\,\phi(x')) = \widetilde{K}(x,x')$ Hence $\varphi(x) = f(x)\,\phi(x)$ gives $\widetilde{K}(x,x') = \varphi(x)\cdot\varphi(x')$.

Submit

You have used 1 of 2 attempts

Answers are displayed within the problem

Kernel Composition Rules 2

0/1 point (graded)

Let x and x' be two vectors of the same dimension. Use the the definition of kernels and the kernel composition rules from the video above to decide which of the following are kernels. (Note that you can use feature vectors $\phi(x)$ that are not polynomial.)

(Choose all those apply.)

- [x] 1
- [x] $x \cdot x'$
- [x] $1 + x \cdot x'$
- [x] $(1 + x \cdot x')^2$
- [ ] $\exp(x + x')$, for x,x' $\in \mathbb{R}$
- [ ] $\min(x, x')$, for x,x' $\in \mathbb{Z}$

✖

Submit You have used 1 of 3 attempts

✖ Incorrect (0/1 point)

Discussion

Hide Discussion

Topic: Unit 2 Nonlinear Classification, Linear regression, Collaborative Filtering (2 weeks):Lecture 6. Nonlinear Classification / 6. Kernel Composition Rules

Add a Post

‹ All Posts

Feature vector for the product kernel

discussion posted a day ago by **ptressel** (Community TA)

For the other composition rules, we saw the feature vector corresponding to the new kernel resulting from the composition, but not for a product of kernels. Hmm, I wonder why... Ok, here's my stab at it. I'm going to inflict this on you, because I balked at going past this without working it out. Please let me know if you see goofs... (Oh, and if you know how to get MathJax to left-align a block of equations, please tell me how! flalign doesn't seem to be recognized.) Spoiler alert: Stop here if you want to try this yourself.

Say that:

$$K_1(x, x') = \phi_1(x) \cdot \phi_1(x') = \sum_{j_1=1}^{d_1} \phi_1^{(j_1)}(x)\,\phi_1^{(j_1)}(x') \quad \text{where} \quad \phi_1(\cdot) \in \mathbb{R}^{d_1}$$

$$K_2(x, x') = \phi_2(x) \cdot \phi_2(x') = \sum_{j_2=1}^{d_2} \phi_2^{(j_2)}(x)\,\phi_2^{(j_2)}(x') \quad \text{where} \quad \phi_2(\cdot) \in \mathbb{R}^{d_2}$$

We want to find a feature vector $\phi(x)$ for the product kernel:

$$K(x, x') = K_1(x, x')\,K_2(x, x') = \phi(x) \cdot \phi(x')$$

So next let's substitute in the definitions of K_1 and K_2.

$$K_1(x, x')\,K_2(x, x') = (\phi_1(x) \cdot \phi_1(x'))\,(\phi_2(x) \cdot \phi_2(x'))$$

$$= \left(\sum_{j_1=1}^{d_1} \phi_1^{(j_1)}(x)\,\phi_1^{(j_1)}(x')\right)\left(\sum_{j_2=1}^{d_2} \phi_2^{(j_2)}(x)\,\phi_2^{(j_2)}(x')\right)$$

$$= \sum_{j_1=1}^{d_1} \sum_{j_2=1}^{d_2} \phi_1^{(j_1)}(x)\,\phi_2^{(j_2)}(x)\,\phi_1^{(j_1)}(x')\,\phi_2^{(j_2)}(x')$$

Hmm, now we've got $\phi_1^{(j_1)}(\cdot)\,\phi_2^{(j_2)}(\cdot)$ first with argument x, then with x'. That looks like almost what we need. There are $d_1 d_2$ terms, indexed by (j_1, j_2). So we just need to line those up and give them a new index j that runs from 1 to $d = d_1 d_2$. That is, we want the components of our combined $\phi(x)$ to be:

$$\phi^{(j)}(x) = \phi_1^{(j_1)}(x)\,\phi_2^{(j_2)}(x)$$

We're essentially done -- we just need to choose how to assign each j to some unique pair of (j_1, j_2) indices. How about if we count up j_1 as the "outer loop" and j_2 as the "inner loop"? i.e.

$$(j_1, j_2) = (1, 1) \to j = 1$$

$$(j_1, j_2) = (1, 2) \to j = 2$$

$$\ldots$$

$$(j_1, j_2) = (1, d_2) \to j = d_2$$

$$(j_1, j_2) = (2, 1) \to j = d_2 + 1$$

$$\ldots$$

$(j_1, j_2) = (d_1, d_2) \rightarrow j = d$

Which is the long way of saying:

$j = (j_1 - 1) d_2 + j_2$

So now we have a feature vector $\phi(x)$ with $d = d_1 d_2$ components, each of which is a product of a component from each of the original feature vectors. Replace the double loop over j_1, j_2 with a loop over j.

$K(x, x') = \sum_{j=1}^{d} \phi^{(j)}(x) \phi^{(j)}(x') = \phi(x) \cdot \phi(x')$

This post is visible to everyone.

0 responses

Preview

Submit