

EdX and its Members use cookies and other tracking technologies for performance, analytics, and marketing purposes. By using this website, you accept this use. Learn more about these technologies in the [Privacy Policy](#).



MITx: 6.86x

Machine Learning with Python-From Linear Models to Deep Learning

[Help](#)[sandipan_dey](#)

[Unit 2 Nonlinear Classification](#),
[Linear regression, Collaborative](#)

[Course](#) > [Filtering \(2 weeks\)](#)

> [Lecture 6. Nonlinear Classification](#) > 6. Kernel Composition Rules

6. Kernel Composition Rules

Kernel Composition Rules



Feature engineering, kernels

Composition rules:

1. $K(x, x') = 1$ is a kernel function. $\phi(x) = 1$
2. Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ and $K(x, x')$ is a kernel. Then so is $\tilde{K}(x, x') = f(x)K(x, x')f(x')$ $\phi(x) = f(x)\phi(x)$
- 3. If $K_1(x, x')$ and $K_2(x, x')$ are kernels, then $K(x, x') = K_1(x, x') + K_2(x, x')$ is a kernel $\phi(x) = \begin{bmatrix} \phi_1(x) \\ \phi_2(x) \end{bmatrix}$
4. If $K_1(x, x')$ and $K_2(x, x')$ are kernels, then $K(x, x') = K_1(x, x')K_2(x, x')$ is a kernel

$$\overline{(x \cdot x')} + \overline{(x \cdot x')^2}$$

$$\phi(x) = x$$

that's called a linear kernel, where ϕ 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.



4:15 / 4:15

Speed 1.50x



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 $f : \mathbb{R}^d \rightarrow \mathbb{R}$ 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 $\widetilde{K}(x, x') = \varphi(x) \cdot \varphi(x')$.

You have used 1 of 2 attempts

i Answers are displayed within the problem

Kernel Composition Rules 2

1/1 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.)

☒ 1 ✓

☒ $x \cdot x'$ ✓

☒ $1 + x \cdot x'$ ✓

☒ $(1 + x \cdot x')^2$ ✓

☒ $\exp(x + x')$, for $x, x' \in \mathbb{R}$ ✓

☐ $\min(x, x')$, for $x, x' \in \mathbb{Z}$

✓

Solution:

We go through the choices in order:

- Yes, for $\phi(x) = 1$.
- Yes, for $\phi(x) = x$.
- Yes, since the sum of kernels are kernels. In this case, we can also easily see $\phi(x) = [1, x]^T$ works.
- Yes, since the product of kernels are kernels. (In this case, factoring the kernel as dot products are more involved, and the composition rule saves this work.)
- Yes, for $\phi(x) = \exp(x)$.
- No. For example, $\min(-1, -1) = -1 < 0$ and hence cannot be written as a dot product and is not a valid kernel.

Submit

You have used 2 of 3 attempts

 Answers are displayed within the problem

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

Kernel Composition Rules 2

discussion posted 4 days ago by Mark B2 (Community TA)

It may be helpful to recall $\exp(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \dots$

This post is visible to everyone.

Add a Response

4 responses

tharit tangkij

4 days ago

Thanks ! Reminding me about this fact helps me understand the lecture on Radial basis Kernel.

Add a comment

nr7116

3 days ago

thanks- I was stuck on that. You nudged me forward

Add a comment

mrBB (Community TA)
2 days ago



It is probably easier to just remember that $\exp(a + b) = \exp(a) \cdot \exp(b)$ I doubt if an infinite sum is (intended to be) relevant here. We can't assume without proof that if $K_n(x, x')$ is a sequence of kernels, it necessarily follows that $\lim_{n \rightarrow \infty} K_n(x, x')$ is kernel as well. (There is plenty of examples where a property that holds for all elements in a sequence, is not a property of its limit.)

I also think that in the model solution, 5th bullet, it should read $f(x)$ instead of $\phi(x)$.



Both $f(x)$ and $\phi(x)$ works.

For $f(x)$, we can treat it as $\exp(x) \cdot 1 \cdot \exp(x')$

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

And "infinite sum" is not needed for this question, but it is helpful for understand next lecture.

posted a day ago by **CoolZ** (Community TA)



The infinite sum indeed made more sense after having watched the next lecture.

posted a day ago by **mrBB** (Community TA)

Add a comment

DataSorcerer
about 6 hours ago



Thanks for the tip. It helped.

Add a comment

Showing all responses

Add a response:

Preview

Submit

Learn About Verified Certificates