Support Vector Machines
Part 3:
The Radial Kernel

Specifically, we're going to talk about the **Radial Kernel's** parameters...

Radial Kernel

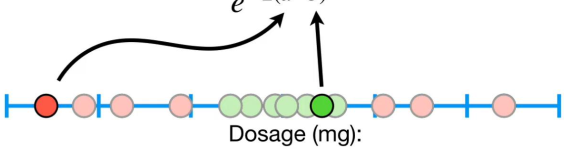
$$e^{-\gamma(a-b)^2}$$

...how the **Radial Kernel** calculates high-dimensional relationships...

Radial Kernel

$$e^{-\gamma(a-b)^2}$$

$$e^{-2(a-b)^2}$$



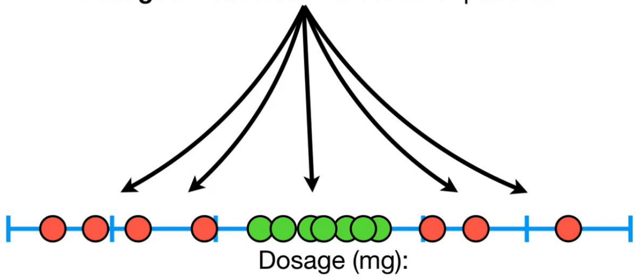
...and then show you how the **Radial Kernel** works in infinite dimensions.

Radial Kernel

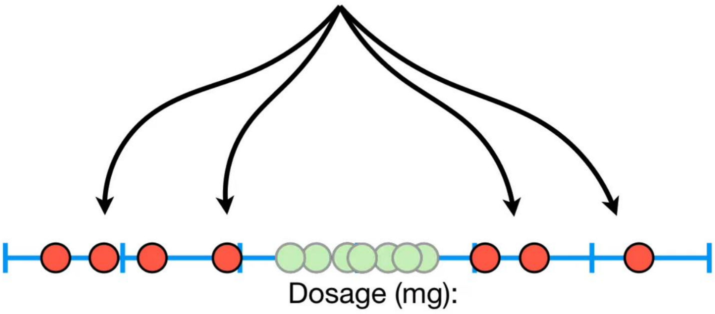
$$e^{-\gamma(a-b)^2}$$

$$e^{-\frac{1}{2}(a-b)^2} = \left(s, s\sqrt{\frac{1}{1!}}a, s\sqrt{\frac{1}{2!}}a^2, \ldots, s\sqrt{\frac{1}{\infty!}}a^{\infty}\right) \cdot \left(s, s\sqrt{\frac{1}{1!}}b, s\sqrt{\frac{1}{2!}}b^2, \ldots, s\sqrt{\frac{1}{\infty!}}b^{\infty}\right)$$

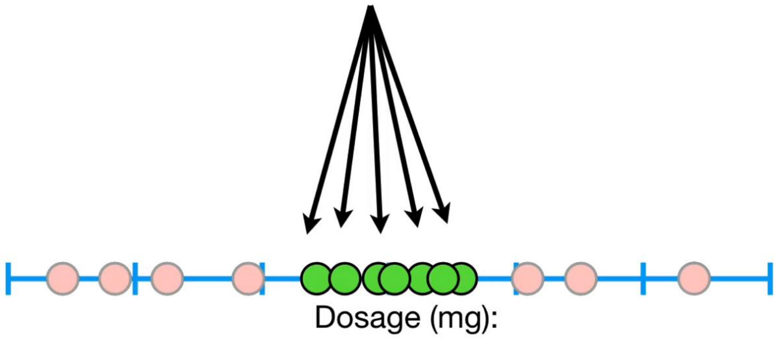
In the **StatQuest** on **Support Vector Machines**, we had a **Training Dataset** based on **Drug Dosages** measured in a bunch of patients.



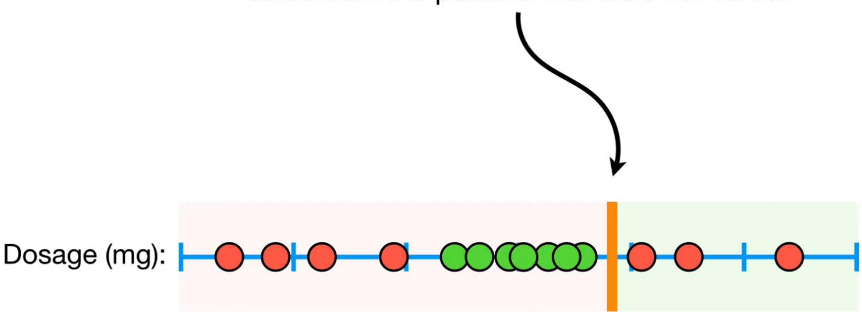
The **red dots** represented patients that were ***not cured***…



…and the **green dots** represented patients that were ***cured***.



Because this **Training Dataset** had so much overlap, we were unable to find a satisfying **Support Vector Classifier** to separate the patients that were cured from the patients that were not cured.



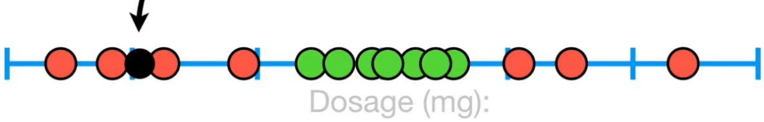
One way to deal with overlapping data is to use a **Support Vector Machine** with a **Radial Kernel**

$$e^{-\gamma(a-b)^2}$$

(aka the **Radial Basis Function**, **RBF**).

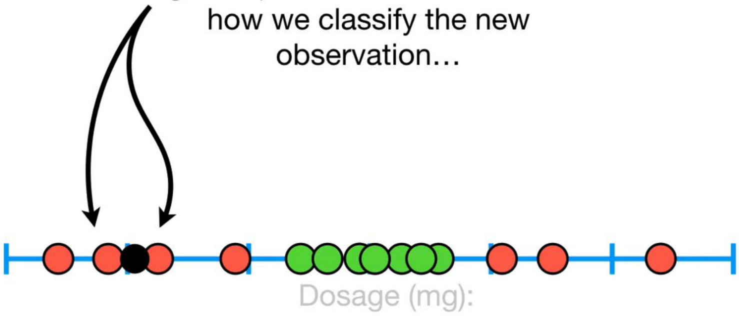
Because the **Radial Kernel** finds **Support Vector Classifiers** in infinite dimensions, it's not possible to visualize what it does.

However, when using it on a new observation like this…

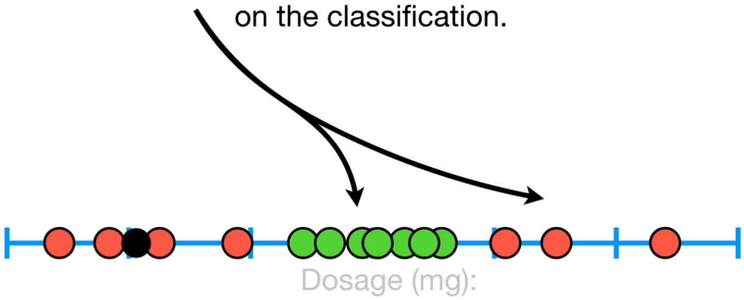


…the **Radial Kernel** behaves like a **Weighted Nearest Neighbor** model.

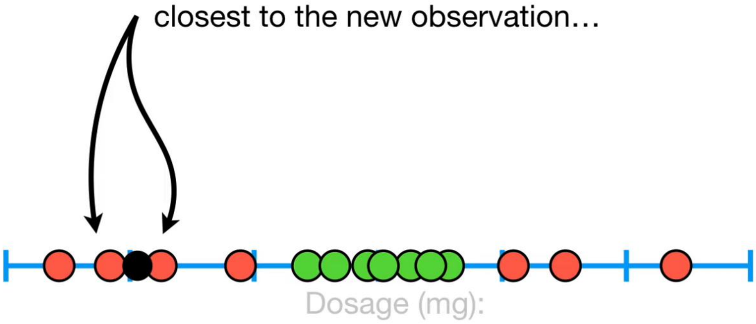
In other words, the closest observations (aka the nearest neighbors) have a lot of influence on how we classify the new observation…



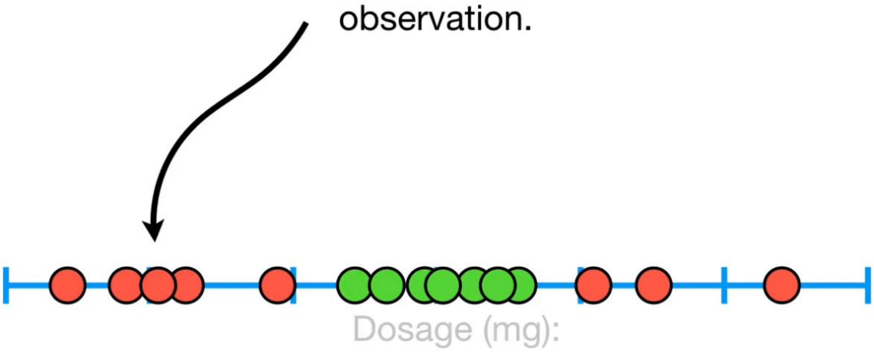
…and observations that are further away have relatively little influence on the classification.



So, since these observations are the closest to the new observation…



…the **Radial Kernel** uses their classification for the new observation.



Now let's talk about how the **Radial Kernel** determines how much influence each observation in the **Training Dataset** has on classifying new observations.

$$e^{-\gamma(a-b)^2}$$

Just like with the **Polynomial Kernel**, ***a*** and ***b*** refer to two different **Dosage** measurements.

$$e^{-\gamma\boxed{(a-b)}^2}$$

The difference between the measurements is then squared, giving us the squared distance between the two observations.

$$e^{-\gamma\boxed{(a-b)^2}}$$

Thus, the amount of influence one observation has on another is a function of the squared distance.

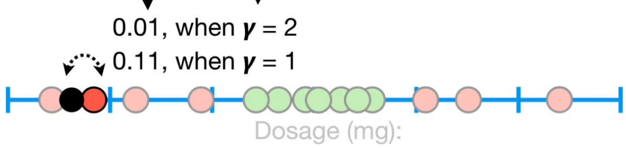
$$e^{-\gamma\boxed{(a-b)^2}}$$

γ (gamma), which is determined by **Cross Validation**, scales the squared distance, and thus, it scales the influence.

$$e^{\boxed{-\gamma}(a-b)^2}$$

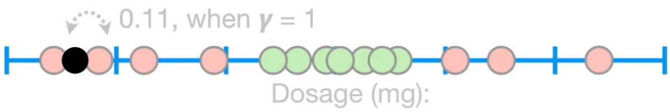
So we see that by scaling the distance, $\boldsymbol{\gamma}$ scales the amount of influence two points have on each other.

$$e^{-\gamma(a-b)^2}$$



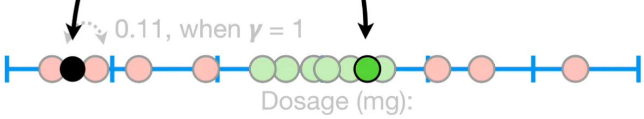
Now let's set $\boldsymbol{\gamma}$ = **1** again…

$$e^{-(a-b)^2}$$



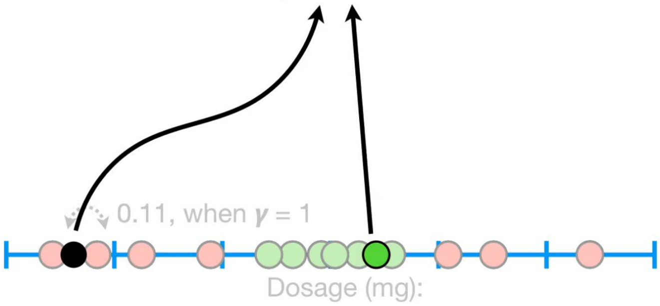
…and determine how much influence two observations have when they relatively far from each other.

$$e^{-(a-b)^2}$$



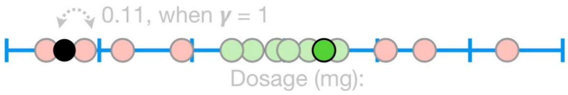
So we plug in the two **Dosages**…

$$e^{-(a-b)^2}$$



…and when the points are relatively far from each other, we get **A Number Very Close to Zero**.

$$e^{-(2.5-16)^2} = e^{-(-13.5)^2} = e^{-182.25} = \boxed{\text{A Number Very Close to Zero.}}$$

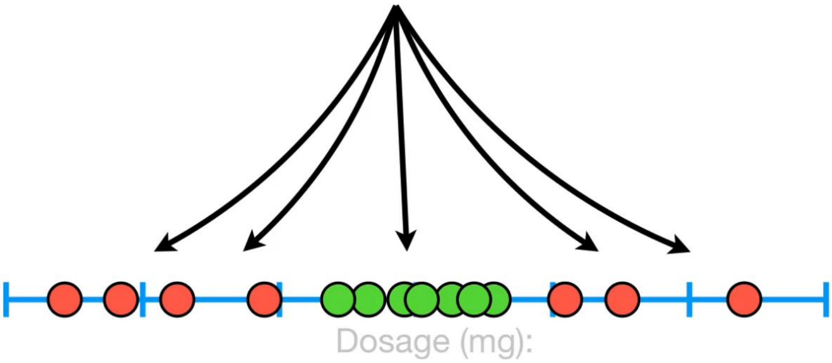


Thus, the further two observations are from each other, the less influence they have on each other.

NOTE: Just like with the **Polynomial Kernel**, when we plug values into the **Radial Kernel**, we get the high-dimensional relationship.

$$e^{-\gamma(a-b)^2} = \text{high-dimensional relationship}$$

Going back to the original
Training Dataset…



a^1b^1

..let's talk about what happens if we take a **Polynomial Kernel** with $\boldsymbol{r}$ = **0** and $\boldsymbol{d}$ = **1**…

$a^1b^1 + a^2b^2$

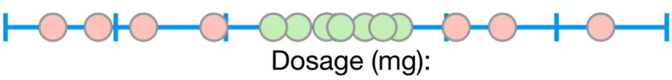
…and add another **Polynomial Kernel** with $\boldsymbol{r}$ = **0** and $\boldsymbol{d}$ = **2**.

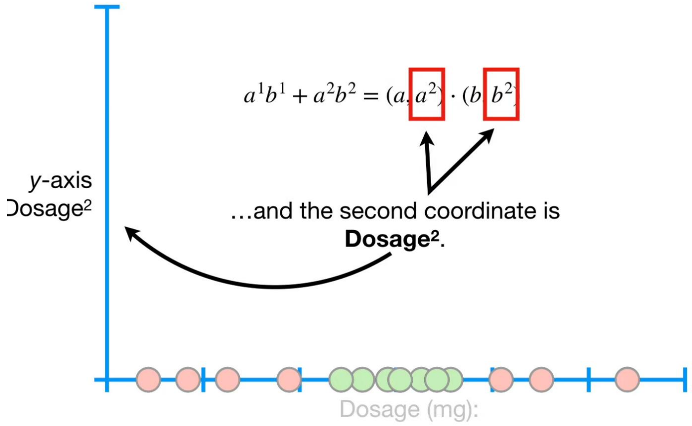
$$a^1b^1 + a^2b^2 = (a, a^2) \cdot (b, b^2)$$

This gives us a **Dot Product** with coordinates for **2-Dimensions**.

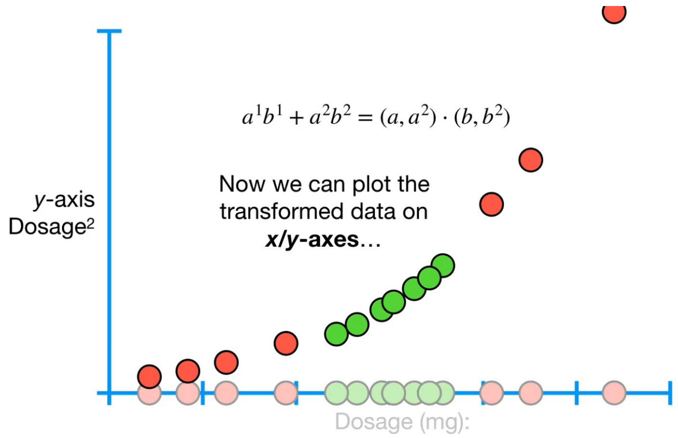
$$a^1b^1 + a^2b^2 = (\boxed{a,}a^2) \cdot (\boxed{b,}b^2)$$

The first coordinate is the original **Dosage**…

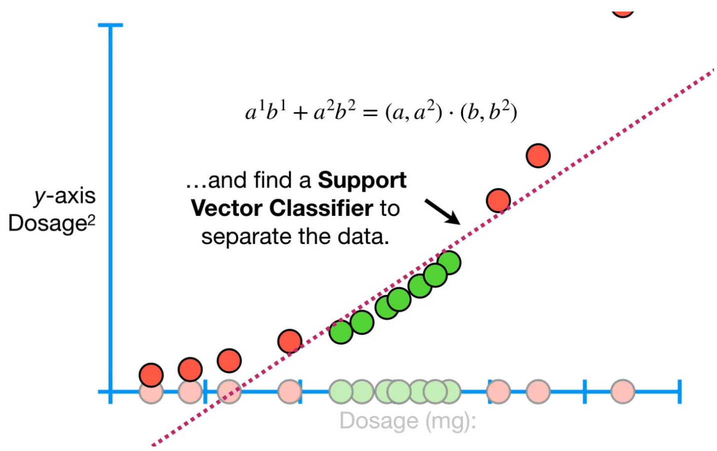




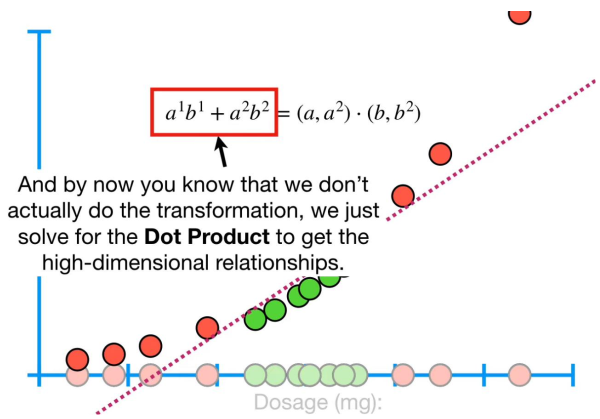
$a^1b^1 + a^2b^2 = (a, \boxed{a^2}) \cdot (b, \boxed{b^2})$
…and the second coordinate is **Dosage²**.
y-axis
Dosage²
Dosage (mg):



$a^1b^1 + a^2b^2 = (a, a^2) \cdot (b, b^2)$
Now we can plot the transformed data on ***x*/*y*-axes**…
y-axis
Dosage²
Dosage (mg):



$a^1b^1 + a^2b^2 = (a, a^2) \cdot (b, b^2)$
…and find a **Support Vector Classifier** to separate the data.
y-axis
Dosage²
Dosage (mg):



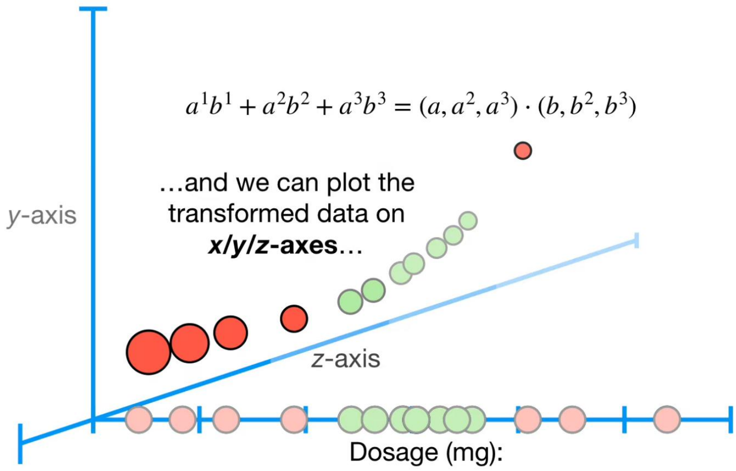
$\boxed{a^1b^1 + a^2b^2} = (a, a^2) \cdot (b, b^2)$
And by now you know that we don't actually do the transformation, we just solve for the **Dot Product** to get the high-dimensional relationships.
Dosage (mg):

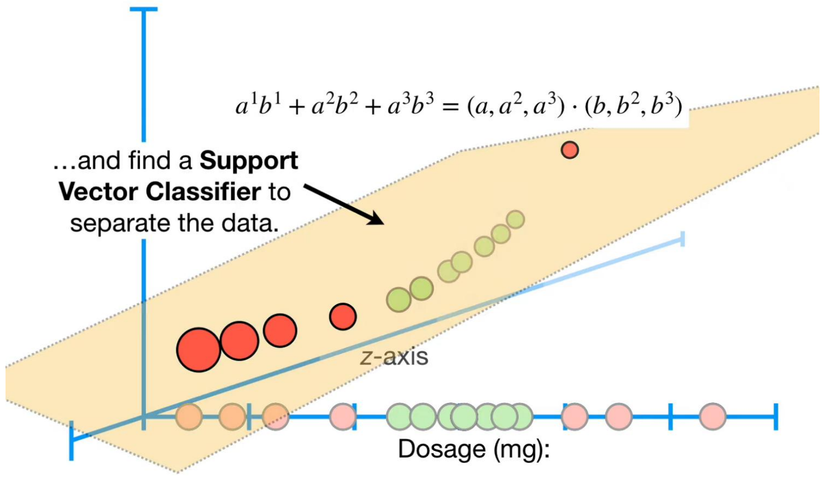
$$a^1b^1 + a^2b^2 + a^3b^3$$

Now, if we added another
Polynomial Kernel with ***r*** = **0**
and ***d*** = **3**…

$$a^1b^1 + a^2b^2 + a^3b^3 = (a, a^2, a^3) \cdot (b, b^2, b^3)$$

…then the **Dot Product** has
coordinates for **3-Dimensions**…





$$a^1b^1 + a^2b^2 + a^3b^3 + \cdots$$

Now, what if we just kept adding **Polynomial Kernels** with ***r* = 0** and increasing ***d*** until ***d* = infinity**?

$$a^1b^1 + a^2b^2 + a^3b^3 + \cdots + a^\infty b^\infty = (a, a^2, a^3, \ldots, a^\infty) \cdot (b, b^2, b^3, \ldots, b^\infty)$$

That would give us a **Dot Product** with coordinates for an *infinite number of dimensions*!!!!

Well, that's exactly what the **Radial Kernel** does, so let's talk about it!!!

Let's start with the **Radial Kernel**…

$$\boxed{e^{-\gamma(a-b)^2}}$$

…and multiply out out the square.

$$e^{-\gamma(a-b)^2} = e^{-\gamma(a^2+b^2-2ab)} = e^{-\gamma(a^2+b^2)}e^{\gamma 2ab}$$

Now, because we can set ***γ*** to anything, let's set it to **1/2**…

$$e^{-\frac{1}{2}(a-b)^2} = e^{-\frac{1}{2}(a^2+b^2-2ab)} = e^{-\frac{1}{2}(a^2+b^2)}e^{ab}$$

…so that this **2** goes away.

Now let's create the **Taylor Series Expansion** of this last term.

$$e^{-\frac{1}{2}(a-b)^2} = e^{-\frac{1}{2}(a^2+b^2-2ab)} = e^{-\frac{1}{2}(a^2+b^2)}\boxed{e^{ab}}$$

This big thing is a **Taylor Series**.

$$\boxed{f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots + \frac{f^{\infty}(a)}{\infty!}(x-a)^{\infty}}$$

Although there are are exceptions, the main idea is that a function, ***f*(*x*)**…

$$\boxed{f(x)} = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots + \frac{f^{\infty}(a)}{\infty!}(x-a)^{\infty}$$

…can be split into an infinite sum.

$$f(x) = \boxed{f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots + \frac{f^{\infty}(a)}{\infty!}(x-a)^{\infty}}$$

Since this is very abstract, let's walk through how we convert e^x into an infinite sum.

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots + \frac{f^{\infty}(a)}{\infty!}(x-a)^{\infty}$$

$$\boxed{e^x}$$

In other words, we are setting
$f(x) = e^x$…

$$\boxed{f(x)} = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots + \frac{f^{\infty}(a)}{\infty!}(x-a)^{\infty}$$

$$\boxed{e^x}$$

…and that means
$f(a) = e^a$.

$$f(x) = \boxed{f(a)} + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots + \frac{f^{\infty}(a)}{\infty!}(x-a)^{\infty}$$

$$e^x = \boxed{e^a}$$

…until we get to infinity.

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots + \boxed{\frac{f^{\infty}(a)}{\infty!}(x-a)^{\infty}}$$

$$e^x = e^a + \frac{e^a}{1!}(x-a) + \frac{e^a}{2!}(x-a)^2 + \frac{e^a}{3!}(x-a)^3 + \cdots + \boxed{\frac{e^a}{\infty!}(x-a)^{\infty}}$$

Thus, this is the **Taylor Series Expansion** of ***e***

$$\boxed{e^x = e^a + \frac{e^a}{1!}(x-a) + \frac{e^a}{2!}(x-a)^2 + \frac{e^a}{3!}(x-a)^3 + \cdots + \frac{e^a}{\infty!}(x-a)^\infty}$$

Now the question is, "What is ***a***?"

etc.

$$e^x = e^a + \frac{e^a}{1!}(x-a) + \frac{e^a}{2!}(x-a)^2 + \frac{e^a}{3!}(x-a)^3 + \cdots + \frac{e^a}{\infty!}(x-a)^\infty$$

The definition of the **Taylor Series** says that ***a*** can be any value as long as ***f***(***a***) exists…

…and since $\mathbf{e^0} = \mathbf{1}$, $\mathbf{e^0}$ exists, so we will set $\boldsymbol{a} = \mathbf{0}$…

…and simplify.

$$e^x = 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 \boxed{+ \cdots + \frac{1}{\infty!}x^\infty}$$

$$e^x = e^0 + \frac{e^0}{1!}(x-0) + \frac{e^0}{2!}(x-0)^2 + \frac{e^0}{3!}(x-0)^3 \boxed{+ \cdots + \frac{e^0}{\infty!}(x-0)^\infty}$$

Going back to the **Radial Kernel**...

$$\boxed{e^{-\frac{1}{2}(a-b)^2}} = e^{-\frac{1}{2}(a^2+b^2-2ab)} = e^{-\frac{1}{2}(a^2+b^2)}e^{ab}$$

$$e^x = 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots + \frac{1}{\infty!}x^\infty$$

...we can now create the **Taylor Series Expansion** of this last term.

$$e^{-\frac{1}{2}(a-b)^2} = e^{-\frac{1}{2}(a^2+b^2-2ab)} = e^{-\frac{1}{2}(a^2+b^2)}\boxed{e^{ab}}$$

To create the **Taylor Series Expansion** of ***e^ab***, we plug in ***ab*** for ***x***.

$$e^{-\frac{1}{2}(a-b)^2} = e^{-\frac{1}{2}(a^2+b^2-2ab)} = e^{-\frac{1}{2}(a^2+b^2)}\boxed{e^{ab}}$$

$$e^x = 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 \boxed{+ \cdots + \frac{1}{\infty!}x^\infty}$$

$$e^{ab} = 1 + \frac{1}{1!}ab + \frac{1}{2!}(ab)^2 + \frac{1}{3!}(ab)^3 \boxed{+ \cdots + \frac{1}{\infty!}(ab)^\infty}$$

Before we move on, let's remember that when we added up a bunch of **Polynomial Kernels** with ***r*** = **0** and ***d*** going from **0** to **infinity**...

$$\boxed{a^0b^0 + a^1b^1 + a^2b^2 + \cdots + a^\infty b^\infty}$$

...we got a **Dot Product** with coordinates for an infinite number of dimensions.

$$a^0b^0 + a^1b^1 + a^2b^2 + \cdots + a^\infty b^\infty \boxed{= (1, a, a^2, \ldots, a^\infty) \cdot (1, b^1, b^2, \ldots, b^\infty)}$$

Thus, each term in this **Taylor Series Expansion** contains a **Polynomial Kernel** with ***r* = 0** and ***d*** going from **0** to **infinity**.

$$a^0b^0 + a^1b^1 + a^2b^2 + \cdots + a^\infty b^\infty = (1,\, a,\, a^2, \ldots, a^\infty) \cdot (1,\, b^1,\, b^2, \ldots, b^\infty)$$

$$\boxed{e^{ab} = 1 + \frac{1}{1!}ab + \frac{1}{2!}(ab)^2 + \frac{1}{3!}(ab)^3 + \cdots + \frac{1}{\infty!}(ab)^\infty}$$

With that in mind, the **Dot Product** for ***e^{ab}*** is…

$$\boxed{e^{ab} = 1 + \frac{1}{1!}ab + \frac{1}{2!}(ab)^2 + \frac{1}{3!}(ab)^3 + \cdots + \frac{1}{\infty!}(ab)^\infty}$$

…this.

$$\boxed{e^{ab} = (1, \sqrt{\frac{1}{1!}}a, \sqrt{\frac{1}{2!}}a^2, \sqrt{\frac{1}{3!}}a^3, \ldots, \sqrt{\frac{1}{\infty!}}a^\infty) \cdot (1, \sqrt{\frac{1}{1!}}b, \sqrt{\frac{1}{2!}}b^2, \sqrt{\frac{1}{3!}}b^3, \ldots, \sqrt{\frac{1}{\infty!}}b^\infty)}$$

Going back to the original **Radial Kernel**…

$$e^{-\frac{1}{2}(a-b)^2} = e^{-\frac{1}{2}(a^2+b^2-2ab)} = e^{-\frac{1}{2}(a^2+b^2)}e^{ab}$$

…we can plug in the **Dot Product** for ***e^{ab}***.

$$e^{-\frac{1}{2}(a-b)^2} = e^{-\frac{1}{2}(a^2+b^2-2ab)} = e^{-\frac{1}{2}(a^2+b^2)}\boxed{e^{ab}}$$

$$\boxed{e^{ab} = (1, \sqrt{\frac{1}{1!}}a, \sqrt{\frac{1}{2!}}a^2, \sqrt{\frac{1}{3!}}a^3, \ldots, \sqrt{\frac{1}{\infty!}}a^\infty) \cdot (1, \sqrt{\frac{1}{1!}}b, \sqrt{\frac{1}{2!}}b^2, \sqrt{\frac{1}{3!}}b^3, \ldots, \sqrt{\frac{1}{\infty!}}b^\infty)}$$

Now the **Radial Kernel** is equal to this term...

$$e^{-\frac{1}{2}(a-b)^2} = e^{-\frac{1}{2}(a^2+b^2-2ab)} = \boxed{e^{-\frac{1}{2}(a^2+b^2)}} e^{ab}$$

$$e^{-\frac{1}{2}(a-b)^2} = \boxed{e^{-\frac{1}{2}(a^2+b^2)}} \left[(1, \sqrt{\frac{1}{1!}}a, \sqrt{\frac{1}{2!}}a^2, \ldots, \sqrt{\frac{1}{\infty!}}a^\infty) \cdot (1, \sqrt{\frac{1}{1!}}b, \sqrt{\frac{1}{2!}}b^2, \ldots, \sqrt{\frac{1}{\infty!}}b^\infty) \right]$$

...times the **Dot Product**.

$$e^{-\frac{1}{2}(a-b)^2} = e^{-\frac{1}{2}(a^2+b^2-2ab)} = e^{-\frac{1}{2}(a^2+b^2)} e^{ab}$$

$$e^{-\frac{1}{2}(a-b)^2} = e^{-\frac{1}{2}(a^2+b^2)} \boxed{\left[(1, \sqrt{\frac{1}{1!}}a, \sqrt{\frac{1}{2!}}a^2, \ldots, \sqrt{\frac{1}{\infty!}}a^\infty) \cdot (1, \sqrt{\frac{1}{1!}}b, \sqrt{\frac{1}{2!}}b^2, \ldots, \sqrt{\frac{1}{\infty!}}b^\infty) \right]}$$

To make the **Radial Kernel** all one **Dot Product** instead of something times a **Dot Product**...

$$e^{-\frac{1}{2}(a-b)^2} = \boxed{e^{-\frac{1}{2}(a^2+b^2)} \left[(1, \sqrt{\frac{1}{1!}}a, \sqrt{\frac{1}{2!}}a^2, \ldots, \sqrt{\frac{1}{\infty!}}a^\infty) \cdot (1, \sqrt{\frac{1}{1!}}b, \sqrt{\frac{1}{2!}}b^2, \ldots, \sqrt{\frac{1}{\infty!}}b^\infty) \right]}$$

...we just multiply both parts of the **Dot Product** by the square root of this term.

$$e^{-\frac{1}{2}(a-b)^2} = \boxed{e^{-\frac{1}{2}(a^2+b^2)}} \left[(1, \sqrt{\frac{1}{1!}}a, \sqrt{\frac{1}{2!}}a^2, \ldots, \sqrt{\frac{1}{\infty!}}a^\infty) \cdot (1, \sqrt{\frac{1}{1!}}b, \sqrt{\frac{1}{2!}}b^2, \ldots, \sqrt{\frac{1}{\infty!}}b^\infty) \right]$$

So we can fit everything on to the screen, let's let ***s*** equal the square root of the first term.

$$\boxed{s = \sqrt{e^{-\frac{1}{2}(a^2+b^2)}}}$$

$$e^{-\frac{1}{2}(a-b)^2} = \boxed{e^{-\frac{1}{2}(a^2+b^2)}} \left[(1, \sqrt{\frac{1}{1!}}a, \sqrt{\frac{1}{2!}}a^2, \ldots, \sqrt{\frac{1}{\infty!}}a^\infty) \cdot (1, \sqrt{\frac{1}{1!}}b, \sqrt{\frac{1}{2!}}b^2, \ldots, \sqrt{\frac{1}{\infty!}}b^\infty) \right]$$

Now we multiply the **Dot Product** by **s**...

$$\boxed{s = \sqrt{e^{-\frac{1}{2}(a^2+b^2)}}}$$

$$e^{-\frac{1}{2}(a-b)^2} = e^{-\frac{1}{2}(a^2+b^2)} \left[(1, \sqrt{\frac{1}{1!}}a, \sqrt{\frac{1}{2!}}a^2, \ldots, \sqrt{\frac{1}{\infty!}}a^\infty) \cdot (1, \sqrt{\frac{1}{1!}}b, \sqrt{\frac{1}{2!}}b^2, \ldots, \sqrt{\frac{1}{\infty!}}b^\infty) \right]$$

$$e^{-\frac{1}{2}(a-b)^2} = (s, s\sqrt{\frac{1}{1!}}a, s\sqrt{\frac{1}{2!}}a^2, \ldots, s\sqrt{\frac{1}{\infty!}}a^\infty) \cdot (s, s\sqrt{\frac{1}{1!}}b, s\sqrt{\frac{1}{2!}}b^2, \ldots, s\sqrt{\frac{1}{\infty!}}b^\infty)$$

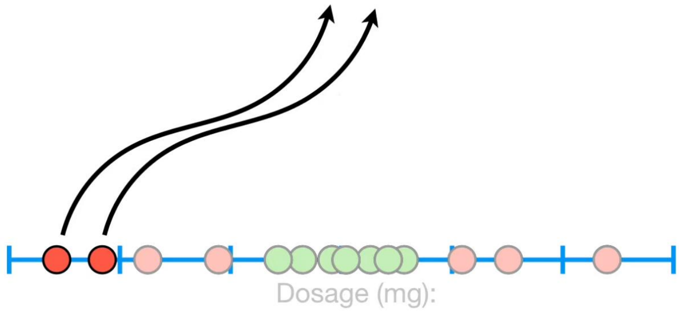
…and, at long last, we see that the **Radial Kernel** is equal to a **Dot Product** that has coordinates for an infinite number of dimensions.

$$e^{-\frac{1}{2}(a-b)^2} = e^{-\frac{1}{2}(a^2+b^2)}\left[(1, \sqrt{\frac{1}{1!}}a, \sqrt{\frac{1}{2!}}a^2, \ldots, \sqrt{\frac{1}{\infty!}}a^{\infty}) \cdot (1, \sqrt{\frac{1}{1!}}b, \sqrt{\frac{1}{2!}}b^2, \ldots, \sqrt{\frac{1}{\infty!}}b^{\infty})\right]$$

$$\boxed{e^{-\frac{1}{2}(a-b)^2}} = (s, s\sqrt{\frac{1}{1!}}a, s\sqrt{\frac{1}{2!}}a^2, \ldots, s\sqrt{\frac{1}{\infty!}}a^{\infty}) \cdot (s, s\sqrt{\frac{1}{1!}}b, s\sqrt{\frac{1}{2!}}b^2, \ldots, s\sqrt{\frac{1}{\infty!}}b^{\infty})$$

That means that when we plug numbers into the **Radial Kernel**…

$$e^{-(2.5-4)^2}$$



…the value we get at the end is the relationship between the two points in **infinite-dimensions**.

$$e^{-(2.5-4)^2} = e^{-(-1.5)^2} = e^{-2.25} = \boxed{0.11}$$