Quiz: Using Kernels

Michael Jayasuriya
University of California, Berkeley
mjayasur@berkeley.edu

Anthony Patitucci
University of California, Berkeley
apatitucci1@berkeley.edu

Kevin Merino
University of California, Berkeley
kmerino@berkeley.edu

Rachel Wang
University of California, Berkeley
rachel.pj.wang@berkeley.edu

Sam Yuen
University of California, Berkeley
yuens@berkeley.edu

Checking kernels

For questions 1 and 2, suppose that K_1 and K_2 are kernels with feature maps ϕ_1 and ϕ_2. For each problem, state whether K is a valid kernel function or not. If they are kernels, give a simple explanation for why. If not, a simple counterexample is sufficient to show why it is not.

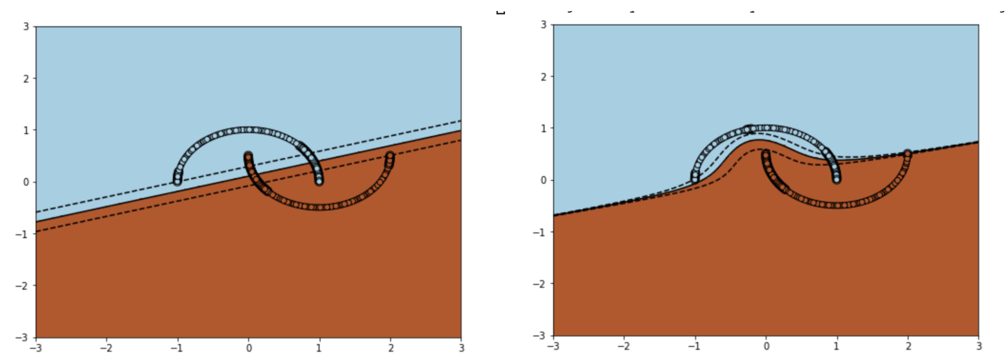
1. $K(x, z) = cK_1(x, z)$ for $c > 0$
   Solution:
   If $K(x, z) = \langle \phi_1(x), \phi_1(z) \rangle$, then we know that $cK(x, z) = \langle \sqrt{c}\phi_1(x), \sqrt{c}\phi_1(z) \rangle$. We know that because $c > 0$, $\sqrt{c} > 0$. Therefore this satisfies the positive semi definite condition. It is symmetric because dot products are commutable.
2. $K(x, z) = cK_1(x, z)$ for $c < 0$ and suppose that there is an x_0 where $K_1(x_0, x_0) > 0$
   Solution:
   Proof by counterexample. Notice that $cK_1(x, z) > 0$ whenever $K_1(x, z) < 0$, because c is always negative. However, because we know that there is an x_0 such that $K(x_0, x_0) > 0$, we know that $K(x_0, x_0) < 0$.
3. $K(x, z) = K_1(x, z) + K_2(x, z)$
   Solution:
   Notice that $K_1(x, z) > 0$ and $K_2(x, z) > 0$. From here, it is quite simple to see that no matter what $K_1(x, z) + K_2(x, z) > 0$. For the kernel function $K_1(x, z)$ let's call the gram matrix K_1, and for the kernel function $K_1(x, z)$ let's call the gram matrix K_2. From here, we can see that the gram matrix for $K(x, z) = K = K_1 + K_2$. The addition of these two positive semi definite matrices will be positive semi definite.
4. $K(x, z) = K_1(x, z) - K_2(x, z)$
   Solution:
   For a counter example, define $K_1(x, z) = x^T z$, and $K_2(x, z) = 2x^T z$. In addition, define $x = [1, 0]$ and $z = [0, 1]$ These two functions have the Gram matrices K_1 and K_2. With the assumptions, $K_1 = I$ and $K_2 = 2I$. If we subtract $K = K_1 - K_2$, we see that $K = -1 * I$, which has all negative eigenvalues and therefore is not positive semi definite.

Identifying kernels

For questions 5 and 6, below are images of decision boundaries created by SVMs using different kernels. For each of the following images, describe what kernel you believe was used to create the boundary, and provide a sentence explaining your decision.

Choose from the following kernels: rbf, linear, sigmoid, and polynomial.

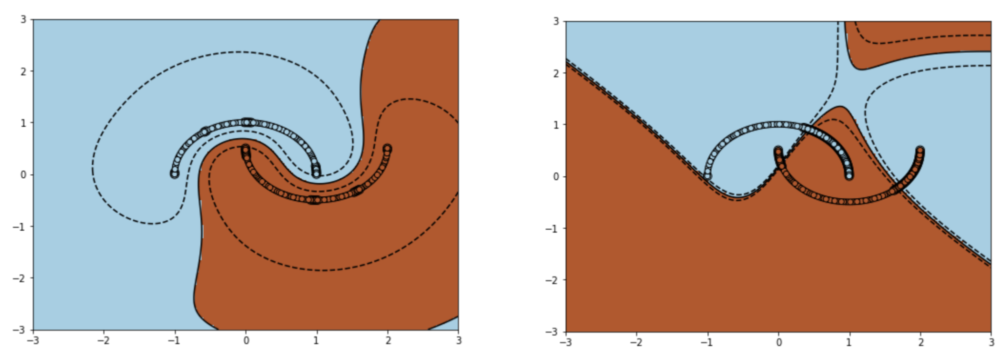
5.



(a) (b)

Solution:
(a) corresponds to the linear kernel. This can be seen from the linear decision boundary.
(b) corresponds to the polynomial kernel. The decision boundary is almost linear with a slight deviation towards the middle. However, it is clear that this decision boundary is not complicated enough to be the rbf or sigmoid kernel, so by process of elimination, it must be the polynomial kernel.

6.



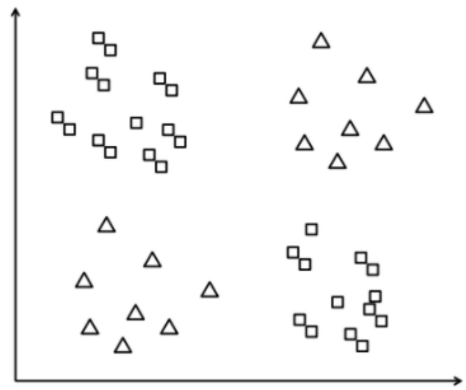
(a) (b)

Solution: Now, we'll deal with the more complicated kernels.
We first look at (b). which corresponds to the sigmoid kernel. We know the sigmoid kernel function is $\tanh(\gamma \cdot X^T y + r)$. The lower decision boundary strongly resembles the tanh function, which and so we arrive at our conclusion.
(a) corresponds to the rbf kernel. The shape of the kernel does not lead us directly to this conclusion. Nevertheless, we can see the rbf kernel greatly outperforms any of the other kernels because the shape of the data is roughly exponential itself, and we know that the rbf kernel is based on an exponential function,

Kernel roulette

For the following scenarios in questions 7 and 8, choose which kernel functions work best.

7. Given the following data samples, which one of the following kernels can we use to separate the two classes? (note: multiple answers may apply)

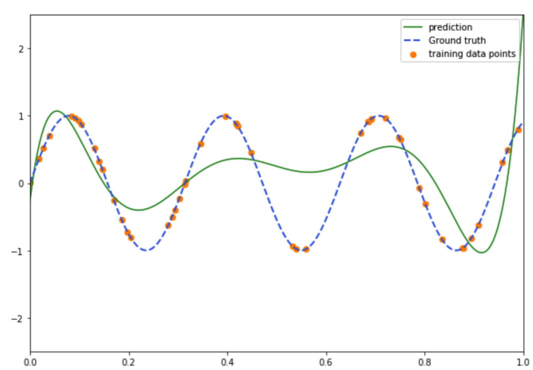


(a)

- ◯ RBF
- ◯ Linear
- ◯ Polynomial
- ◯ None of the above

Solution: RBF and Polynomial

8. Given the following visualization, which was used, a Laplacian kernel or a RBF kernel?



(a)

- ◯ Laplacian
- ◯ RBF

Solution: Laplacian

Building custom kernels

9. For this problem, we will be building a custom kernel with sklearn to classify some data. Fill in the blanks in the following code to make a custom ANOVA kernel. For this problem, keep $d = 3$ and $n = 100$ and $\sigma = 3$ as a fixed value.

$$k(x, z) = \Sigma_{k=1}^{n} exp((-\sigma(x^k - z^k)^2)^d) \tag{1}$$

```
def anova(x, z):
    total = _________
```

```
    for k in range(1, 11):
        total += ________
    return _______
```

Solution:

```
def anova(x, z):
    total = 0
    for k in range(1, 11):
        total += np.exp(-3 * (x**k - z**k) ** 3)
    return total
```

10. Now that we have our custom kernal, fill in the following lines to fit an SVM with our kernal, and use it to fit our data.

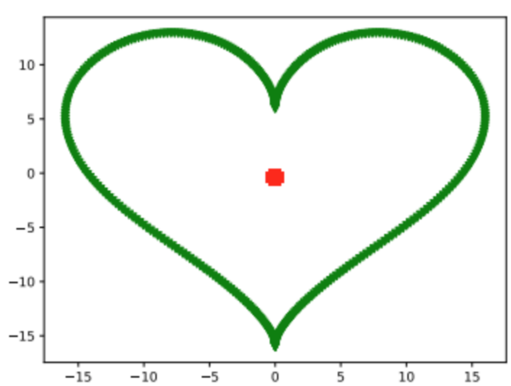
```
from sklearn import svm
# Assume you have some data X, Y
# Create an instance of SVM and fit out data.
clf = svm.SVC(_____)
clf.fit(_____, _____)
```

Solution:

```
from sklearn import svm
# Assume you have some data X, Y
# Create an instance of SVM and fit out data.
clf = svm.SVC(kernal=anova)
clf.fit(X,Y)
```

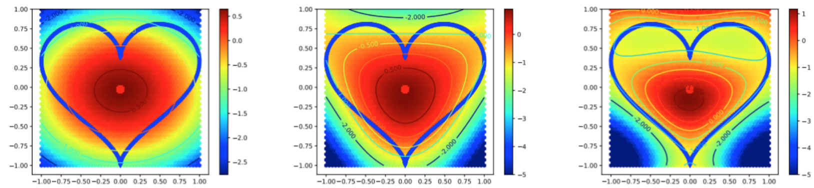
Identifying Hyperparameters

For this problem, let's take a look at this dataset.

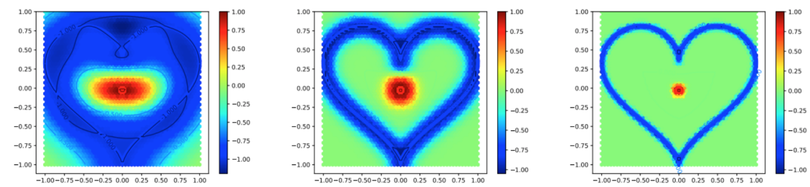


(a)

This is a labeled dataset where the green dots represent datapoints with the label -1 and red dots represent data with the label +1. (The heart looks continuous but is made up of a ton of datapoints) Now, assume that we trained a kernel ridge regression model on this dataset with 6 different RBF kernel's, each with different gammas.



(a)



(b)

These heatmaps are a little confusing, but they are color coded by how the data was classified. For instance, if a point falls in a region that is very red, it will classified by Kernel Ridge Regression to be of class +1. If a point falls in a region that that is very blue, it will be classified as -1 by the Kernel Ridge model. The intensity of the color tells us how positive or negative the Kernel Ridge model's guess was, meaning that the regions in green are classified as not +1 or -1.
Answer the following questions about the heatmaps.

10. Which heatmap was produced by the kernel with the highest gamma? Why?
    Solution:
    The highest gamma is the bottom right, because when a RBF kernel has high gamma it will drive the kernel function to zero when two data points are not equal. In other words, points every point is far away in the eyes of a high gamma kernel.

11. Which heatmap was produced by the kernel with the lowest gamma? Why?
    Solution:
    The lowest gamma is the top left, because when a RBF kernel has low gamma it will drive the kernel function to infinity, even when two data points are not equal. In other words, points every point is very close in the eyes of a low gamma kernel.

References

[1] Kamalika Chaudhuri, University of California, San Diego, CS 151A https://cseweb.ucsd.edu/classes/sp19/cse151-a/hw5sol.pdf [2] Ben Recht et al, University of California, Berkeley, CS 189 https://tbp.berkeley.edu/exams/6371/download/
[3] djs, Stack Overflow User https://stats.stackexchange.com/questions/270967/proof-of-sum-of-kernels-of-concatenated-vector [4] Anant Sahai et al, University of California, Berkeley, CS 189 Homework 4 https://www.eecs189.org/
[5] SKLearn Official Page https://scikit-learn.org/stable/auto$_{e}xamples/svm/plot_{c}ustom_{k}ernel.html$