1 Solution Kernels: Exercise 1.1

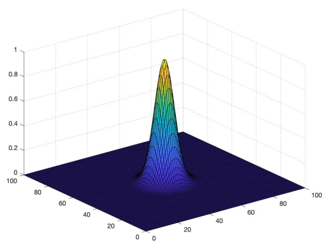
Using the RBF kernel, draw the isolines for one data point x^1. Discuss the effect of modifying the kernel width.

Recall that the equation of the Gaussian (RBF) kernel is:

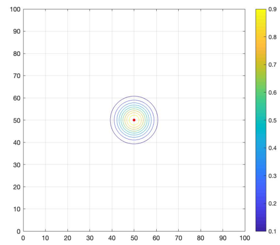
$$k(x, x') = e^{\frac{||x-x'||^2}{2\sigma^2}}, \tag{1}$$

where $||x - x'||$ is the standard Euclidean norm and σ is the kernel width, its hyperparameter.

Fig. 1 shows the surface and contour representation of the function $f(x) = k(x, x_1)$, where $x_1 = (50, 50)$.



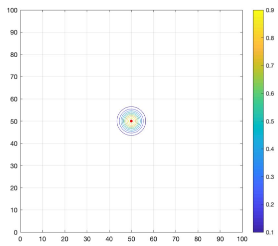
(a)



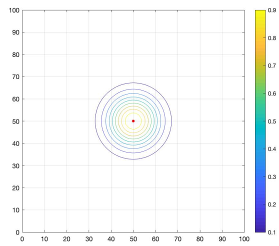
(b)

Figure 1: Gaussian (RBF) Kernel with $\sigma = 5$: (a) surface; (b) contour.

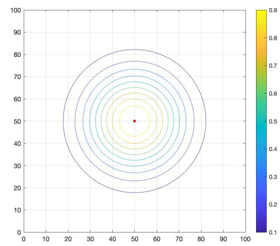
Changing the kernel width of the RBF influences the velocity with which the function decays towards zero. Fig. 2 shows the isoline of the function $f(x) = k(x, x_1)$ for different kernel width.



(a)



(b)



(c)

Figure 2: Comparison of Gaussian (RBF) Kernel with different width: (a)$\sigma = 3$; (b) $\sigma = 8$; (c) $\sigma = 15$.

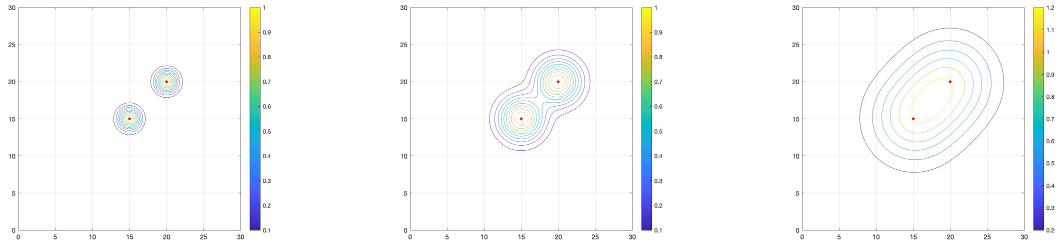
Solution Kernels: Exercise 1.2

Using the RBF kernel, draw the isolines when adding or substracting the each datapoints, namely:

- *Find all* x, *s.t.* $k(x, x^1) + k(x, x^2) =$*cst.*
- *Find all* x, *s.t.* $k(x, x^1) - k(x, x^2) =$*cst.*

We consider two kernels respectively centered in $x_1 = (15, 15)$ and $x_1 = (20, 20)$. For different width of both kernels, Fig. 3 shows the isolines of function

$$f(x) = k(x, x_1) + k(x, x_2). \tag{2}$$



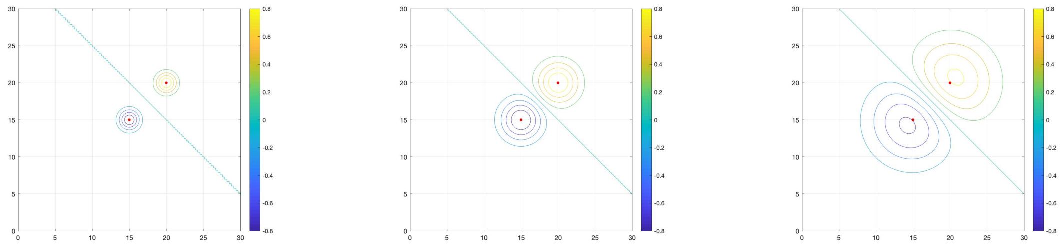
(a) (b) (c)

Figure 3: Sum of two Gaussian (RBF) Kernel with different width: (a)$\sigma = 1$; (b) $\sigma = 2$; (c) $\sigma = 4$.

The two points are regrouped through the isolines. When moving far from the datapoints, the isolines become a tight circle, which englobes the datapoints. Observe that such a construct can cluster the datapoint in one group delineated by one particular value of the associated isoline. This is at the basis of Support Vector Clustering which we will see in a few lectures.

For different width of both kernels, Fig. 4 shows the isolines of function

$$f(x) = k(x, x_1) - k(x, x_2). \tag{3}$$



(a) (b) (c)

Figure 4: Subtraction of two Gaussian (RBF) Kernel with different width: (a)$\sigma = 1$; (b) $\sigma = 2$; (c) $\sigma = 4$.

The two points are well separated with positive and negative values on the isolines. Applying a threshold on the isoline zero allows to separate the two points. This is at the basis of classification when using RBF kernel in Support Vector Machine, whereby one takes the sign of the isoline to determine the label.

Solution Kernels: Exercise 1.3

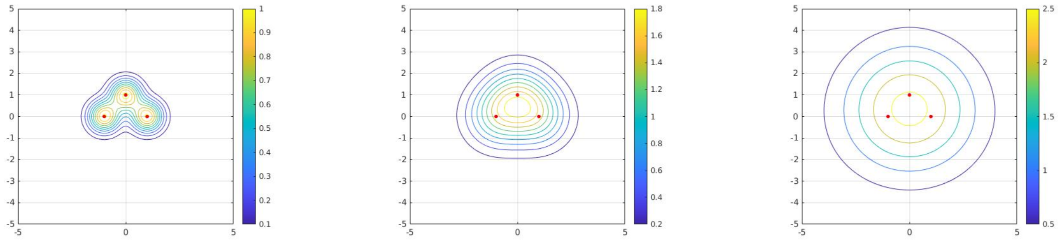
Using the RBF kernel, draw the isolines when combining three data-points, namely:

- *Find all x, s.t. $k(x, x^1) + k(x, x^2) + k(x, x^3) =$cst.*
- *Find all x, s.t. $k(x, x^1) + k(x, x^2) - k(x, x^3) =$cst.*

Observe again that you can either regroup all points when using the additive term or separate one of the points when subtracting its contribution. Consider three RBF kernels at $x_1 = (-1, 0)$, $x_2 = (1, 0)$ and $x_3 = (0, 1)$. Fig. 5 shows, for different length of the kernels, the isolines of the function

$$f(x) = k(x, x_1) + k(x, x_2) + k(x, x_3). \tag{4}$$

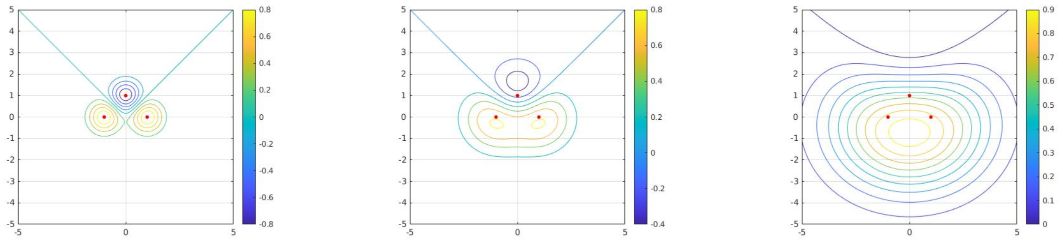
Fig. 6 shows, for different length of the kernels, the isolines of the function



(a) (b) (c)

Figure 5: Sum of three RBF kernels with (a) $\sigma = 0.5$; (b) $\sigma = 1$; (c) $\sigma = 2$.

$$f(x) = k(x, x_1) + k(x, x_2) - k(x, x_3). \tag{5}$$



(a) (b) (c)

Figure 6: Sum and subtraction of three RBF kernels with (a) $\sigma = 0.5$; (b) $\sigma = 1$; (c) $\sigma = 2$.

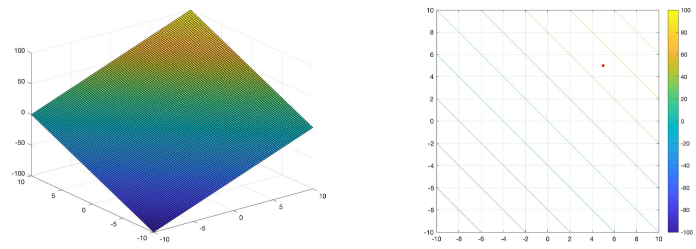
2 Solution Kernels: Exercise 2.1

Using the homogeneous polynomial kernel, draw the isolines as in previous exercise for one data point and two data-points. Discuss the effect of the hyperparametr p on the isolines.

The equation of the homogenous polynomial kernel is:

$$k(x, x') =< x, x' >^p, \tag{6}$$

where $< x, x' >$ is the standard Euclidean dot product and p is the degree of the polynomial.

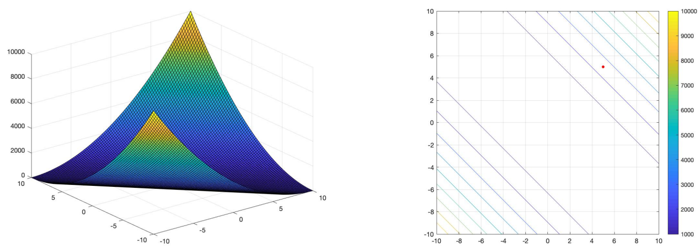


(a) (b)

Figure 7: Homogenous Polynomial Kernel $p = 1$: (a) surface; (b) contour.

The homogeneous polynomial kernel is the equation of a projection. The isolines are orthogonal to the vector pointing from the origin to x^1. They are positives for all x pointing in the same direction as x^1 and negatives otherwise. This is due to the sign of the angle between the two vectors.

We consider as reference point $x_1 = (5, 5)$. Fig. 7 shows the first order polynomial $k(x, x_1)$. Fig. 8 shows the second order polynomial $k(x, x_1)$.



(a) (b)

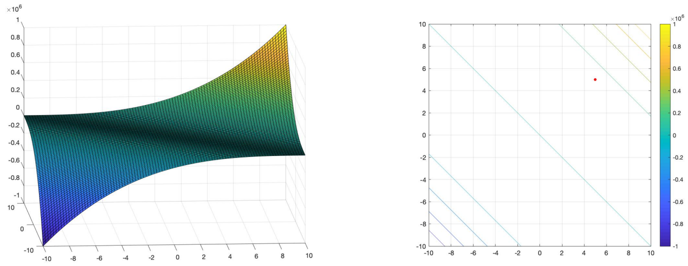
Figure 8: Homogenous Polynomial Kernel $p = 2$: (a) surface; (b) contour.

When we elevate the kernel to the power of 2, the isolines are now all positive. Indeed, we have $k(x, x') =< x, x' >^p= ||x||^2||x^1||^2 \cos(\theta)^2$. The cosine term is now always positive. The isolines are no longer equidistant. The distance decreases with the square of the distance.

When we elevate the kernel to the power of 3, the isolines are again positive and negative. The distance decreases with the cube of the distance, see Fig. 9.

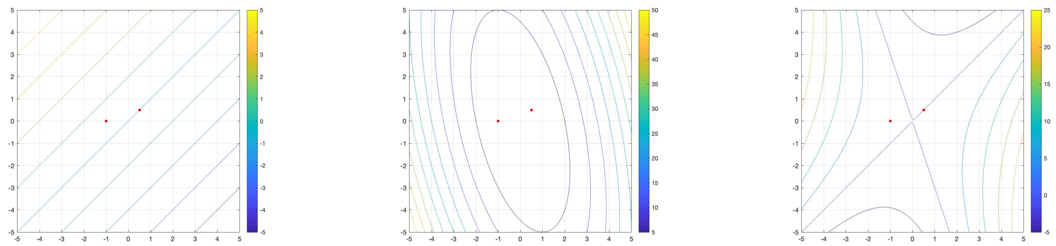
For power larger than 3, we observe the same behavior. The isolines remain lines perpendicular to the datapoints and change sign only for odd number of p.

We consider now the sum and subtraction of two polynomial kernels having as reference points $x_1 = (-1, 0)$ and $x_2 = (0.5, 0.5)$, respectively. Fig.10 shows the isolines for the sum and



(a) (b)

Figure 9: Homogenous Polynomial Kernel $p = 3$: (a) surface; (b) contour.



(a) (b) (c)

Figure 10: (a) Sum of two polynomial kernels with $p = 1$; (b) Sum of two polynomial kernels with $p = 2$; (c) Subtraction of two polynomial kernels with $p = 2$.

subtraction of polynomial kernels with different degrees. In the first case (a) two polynomials kernel of degree 1 are computed as the following sum:

$$f(x) = (x^1)^T x + (x^2)^T x. \tag{7}$$

In the second case (b) two polynomials of second order are added up:

$$f(x) = ((x^1)^T x)^2 + ((x^2)^T x)^2, \tag{8}$$

Calculating the isolines such a functions correspond to find the set of points where the function yields the same value, that is:

$$((x^1)^T x)^2 + ((x^2)^T x)^2 = k, \tag{9}$$

where k is a constant scalar.

If x is 2-dimensions, its coordinates are $x = [x_1, x_2]^T$. When we expand the previous equation, we obtain $ax_1^2 + bx_2^2 + cx_1x_2$. This is the equation of an **ellipse**. The coefficients a, b, c determine the axes of the ellipses and are given by the coordinates of the datapoints x^1 and x^2, such that we have $a = (x_1^1)^2 + (x_1^2)^2$, $b = (x_2^1)^2 + (x_2^2)^2$ and $c = 4x_1^1 x_1^2 x_2^1 x_2^2$. When the datapoints are opposite to one another with respect to the origin, the term c is zero. In this case, the ellipse is centered on the origin and its axis are aligned with the main axes of the frame of reference. Otherwise, the ellipse is centered at the origin but tilted.

In the third case (c) two polynomial of second order are subtracted:

$$f(x) = ((x^1)^T x)^2 - ((x^2)^T x)^2 \tag{10}$$

Calculating the isoline of such a function corresponds to finding the set of point where the function yields the same value:

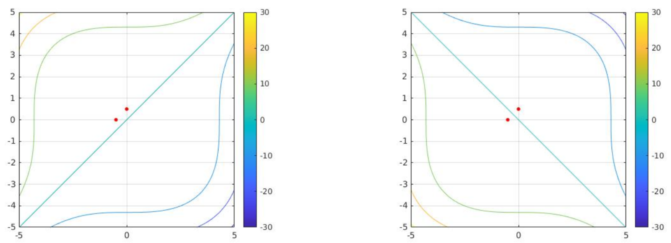
$$((x^1)^T x)^2 - ((x^2)^T x)^2 = k, \tag{11}$$

where k is a constant scalar. This equation reveals that, for $x_1 \neq x_2$ the isolines (level sets) are **hyperbolas**.

When we expand the previous equation, we obtain $ax_1^2 - bx_2^2 + cx_1x_2$. This is the equation of an **hyperbola**. The coefficients a, b, c determine the direction of the hyperbola. It is also centered at the origin and tilts depending on the location of the two points.

When we elevate to power $p = 3$ and $p = 4$, we get deformed ellipses and hyperbolas, still centered at the origin. This is due to the fact that the main shape of the equation entailed in the terms of the form $(ax_1^2 \pm bx_2^2 + cx_1x_2)$ is preserved, with a, b, c three coefficients that depend on the location of the points. This main term is elevated to the power $p - 2$, i.e. we have $((ax_1^2 \pm bx_2^2 + cx_1x_2)^2)^{p-2}$. The terms of higher order have the largest influence.

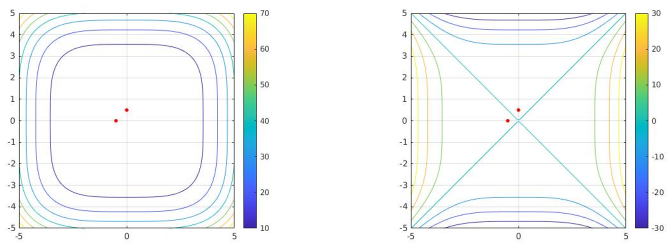
Fig. 11 shows the sum and the subtraction of two polynomial kernels of degree three.



(a) (b)

Figure 11: (a) Sum of two polynomial kernels with $p = 3$; (b) Subtraction of two polynomial kernels with $p = 3$.

Fig. 12 shows the sum and the subtraction of two polynomial kernels of degree four.



(a) (b)

Figure 12: (a) Sum of two polynomial kernels with $p = 4$; (b) Subtraction of two polynomial kernels with $p = 4$.

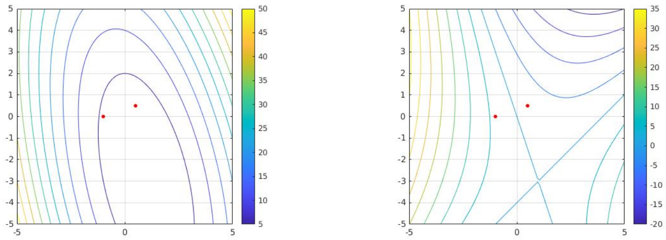
3 Solution Kernels: Exercise 2.2

The equation of the inhomogeneous polynomial kernel is:

$$k(x, x') = (< x, x' > +c)^p, \tag{12}$$

where $< x, x' >$ is the standard Euclidean dot product, p is the degree of the polynomial and c is a constant scalar.

The constant c has two effects: it shifts the origin of the system and acts as a multiplicative factor on the terms of lower orders. The latter increases the value of the isolines. With odd order of the polynomial, we again have negative terms due to the sign of cos. $((ax_1^2 \pm bx_2^2 + cx_1x_2)^2)$ Fig. 13 shows the sum and the subtraction of two inhomogeneous polynomial kernels of degree three.



(a) (b)

Figure 13: (a) Sum of two inhomogeneous polynomial kernels with $p = 2$ and $c = 1$; (b) Subtraction of two inhomogeneous polynomial kernels with $p = 2$ and $c = 1$.

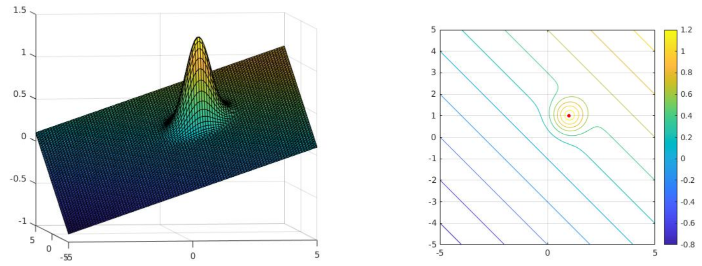
4 Solution Kernels: Exercise 3

Valid kernels can be constructed from the addition and multiplication of kernels. We here illustrated what type of complex shapes we can create through such combinations.

For the function

$$f(x) = k_{poly}(x, x_1) + k_{RBF}(x, x_2), \tag{13}$$

where k_{poly} is a homogenous polynomial kernel and k_{RBF} is a Gaussian (RBF) kernel, Fig 14 shows the surface and the isolines when $x_1 = (1, 1)$ and $x_2 = (1, 1)$.



(a) (b)

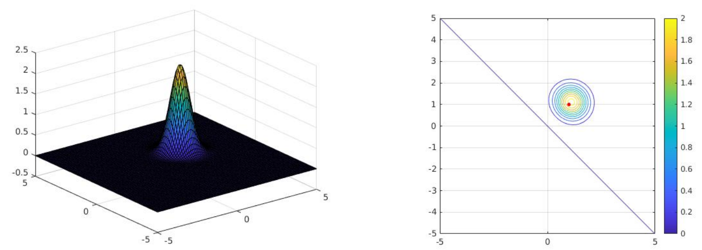
Figure 14: Sum of a Homogenous Polynomial($p = 1$) kernel at $x_1 = (1, 1)$ and RBF ($\sigma = 0.5$) kernels at $x_2 = (1, 1)$.

The RBF allows to generate small bumps in space, while the increasing or decreasing trend of the polynomial is preserved throughout space. As the polynomial kernel yields in general large value whereas the RBF kernel can generate only isolines at maximum 1, it is often necessary to add a multiplicative factor to the RBF to ensure that this term will have an influence.

For the function

$$f(x) = k_{poly}(x, x_1) \times k_{RBF}(x, x_2), \tag{14}$$

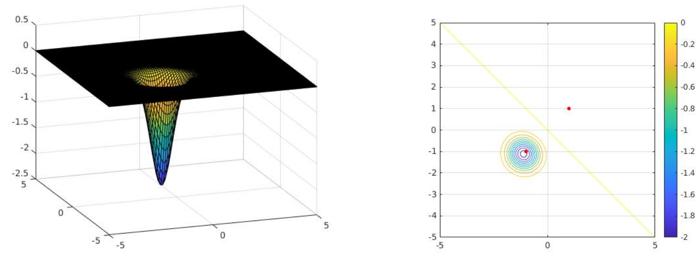
where k_{poly} is a homogenous polynomial kernel and k_{RBF} is a Gaussian (RBF) kernel, Fig 15 shows the surface and the isolines when $x_1 = (1, 1)$ and $x_2 = (1, 1)$. Fig 16 shows the surface



(a) (b)

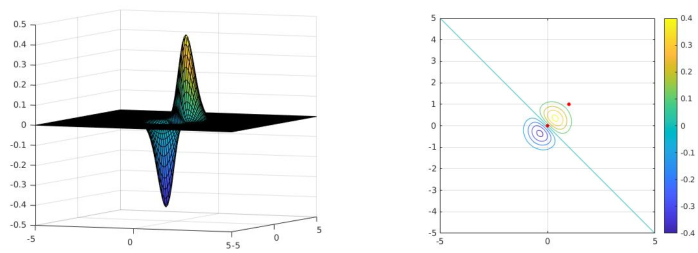
Figure 15: Product of a Homogenous Polynomial($p = 1$) kernel at $x_1 = (1, 1)$ and RBF ($\sigma = 0.5$) kernels at $x_2 = (1, 1)$.

and the isolines when $x_1 = (1, 1)$ and $x_2 = (-1, -1)$. Fig 17 shows the surface and the isolines when $x_1 = (1, 1)$ and $x_2 = (0, 0)$.



(a) (b)

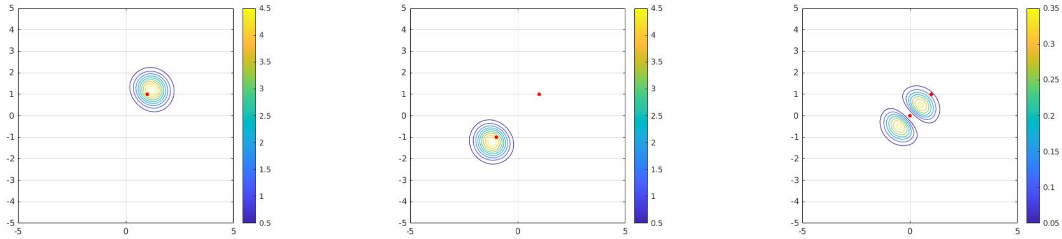
Figure 16: Product of a Homogenous Polynomial($p = 1$) kernel at $x_1 = (1, 1)$ and RBF ($\sigma = 0.5$) kernels at $x_2 = (-1, -1)$.



(a) (b)

Figure 17: Product of a Homogenous Polynomial($p = 1$) kernel at $x_1 = (1, 1)$ and RBF ($\sigma = 0.5$) kernels at $x_2 = (0, 0)$.

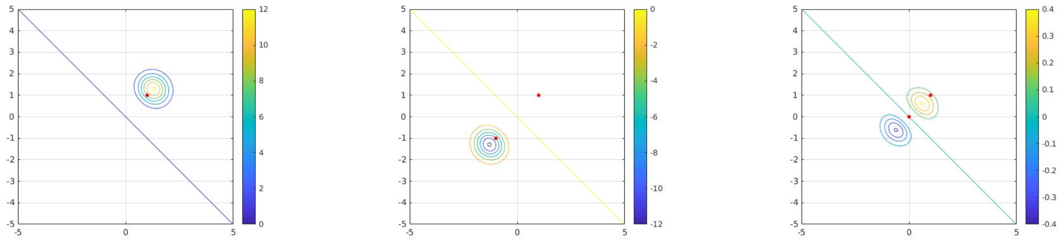
Fig. 18 shows the isolines of the functions given by the product of a second order homogeneous polynomial kernel and a Gaussian (RBF) kernel for three different locations of the latter. Fig. 19



(a) (b) (c)

Figure 18: Product of a Homogenous Polynomial($p = 2$) kernel and RBF ($\sigma = 0.5$) kernel respectively at (a) $x_1 = (1, 1)$ and $x_2 = (1, 1)$; (b) $x_1 = (1, 1)$ and $x_2 = (-1, -1)$; (c) $x_1 = (1, 1)$ and $x_2 = (0, 0)$.

shows the isolines of the functions given by the product of a third order homogeneous polynomial kernel and a Gaussian (RBF) kernel for three different locations of the latter. The sign of



(a) (b) (c)

Figure 19: Product of a Homogenous Polynomial($p = 3$) kernel and RBF ($\sigma = 0.5$) kernel respectively at (a) $x_1 = (1, 1)$ and $x_2 = (1, 1)$; (b) $x_1 = (1, 1)$ and $x_2 = (-1, -1)$; (c) $x_1 = (1, 1)$ and $x_2 = (0, 0)$.

the Gaussian (RBF) kernel is preserved whenever we have multiplication by an even order homogeneous polynomial kernel (the kernel remains positive in all the space). Multiplication by an odd homogeneous polynomial kernel changes the sign of the Gaussian kernel depending on its location. In the examples shown for first and third degree of the homogeneous polynomial the RBF kernel yields positive values if located in the first quadrant negative values if located in the third quadrant.