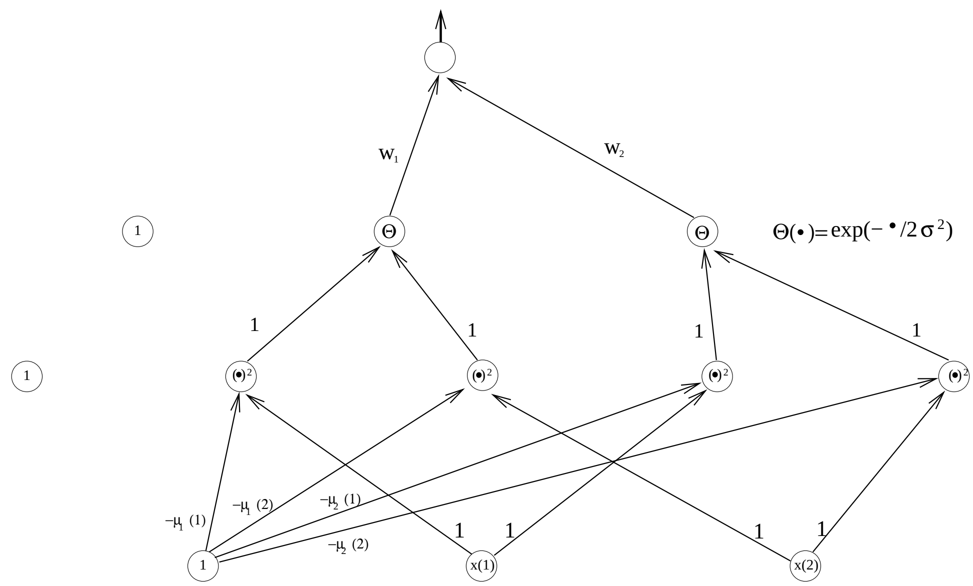
ASSIGNMENT 7, Solutions

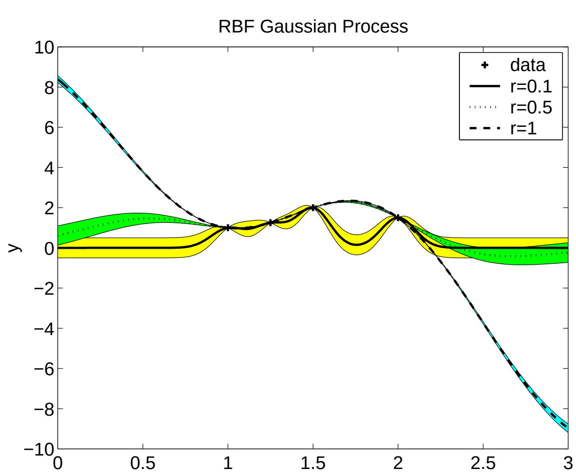
1 (200pts) RBF's and Neural Networks



The first hidden layer serves to compute $\| \mathbf{x} - \mu_i \|^2$ and the second hidden layer computes the exponential kernel. All connection weights not shown are assumed to be zero.

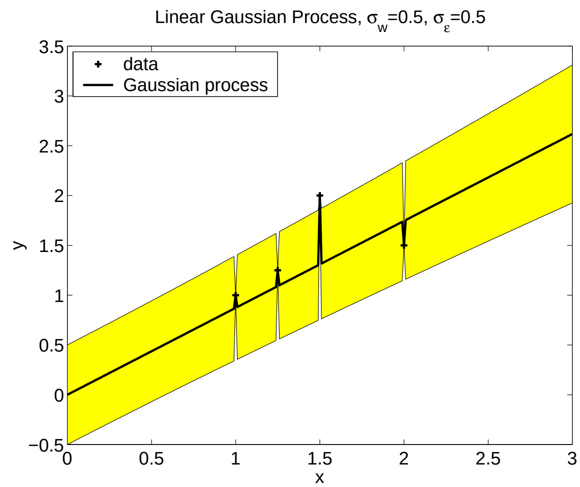
2 (500pts) Gaussian Processes

(a)



The curves show the Gaussian Process fits. The bands illustrate the error bars. Notice that the error bars begome larger as one moves away from the data set. We observe that as r decreases, the curve becomes less smooth which corresponds to decreasing the width of the RBF. This is as expected.

(b)

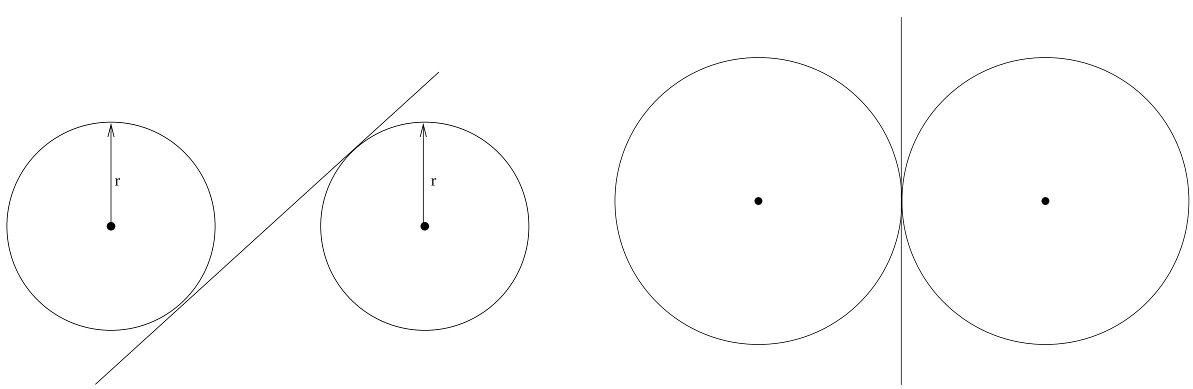


The Gaussian curve is strikingly linear as is to be expected given that we are simulating a linear function.

Notice that in both cases, the Gaussian process fits the data set exactly. It can be shown that this is generally so.

3 (300pts) Support Vector Machines

(a) We give a geometrical argument.



Sub Optimal Optimal

Imagine drawing spheres of a given radius, r, with each data point as center. We know that the optimal hyperplane will be the same distance from each point so the optimal hyperplane will be tangent to both spheres for some radius. A sub-optimal such arrangement is shown in the first figure. It is clear that the two spheres cannot intersect as there is then no separating hyperplane that is tangent to both spheres. If the two spheres cannot intersect, then the largest that r can be is half the distance between the two points. Such a situation is shown in the figure. The only separating hyperplane that is tangent to both spheres is also shown and thus this must be the optimal hyperplane. Hence, the optimal separating hyperplane is the perpendicular bisecting plane. In two dimensions, this is the perpendicular bisector.

In our case, the perpendicular bisector is given by the equation

$$\boxed{x_1 = 0}$$

(b)

i. The data points are unchanged, i.e.,

$$\boxed{\mathbf{z}_1 = (1, 0) \qquad \mathbf{z}_2 = (-1, 0)}$$

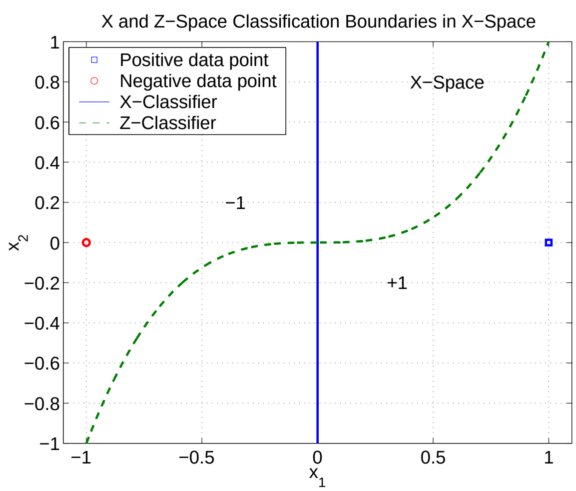
ii. The classifier in Z-space is exactly equivalent to the one in X-space, i.e.,

$$\boxed{z_1 = 0}$$

(c) The X-space classifier has classification boundary $x_1 = 0$. The Z-space classifier has classification boundary $z_1 = 0$, which translates to the classification boundary

$$\boxed{x_1^3 - x_2 = 0}$$

in X-space. These two classification boundaries are shown in the following figure



(d)

$$\boxed{K(\mathbf{x}, \mathbf{y}) = \mathbf{z}(\mathbf{x}) \cdot \mathbf{z}(\mathbf{y}) = (x_1^3 - x_2)(y_1^3 - y_2) + x_1 x_2 y_1 y_2}$$