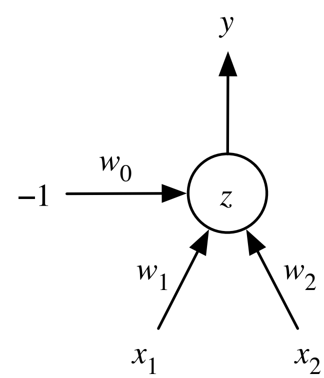
1 Classification

For this problem, we will consider the simple 1-layer neural network shown below, but rather than using a sigmoid function, as we've seen in class, it will use a Gaussian. That is, the output of the final unit, y, is computed as follows:

$$y = g(z)$$
$$g(z) = e^{-z^2}$$
$$z = -w_0 + w_1 x_1 + w_2 x_2$$

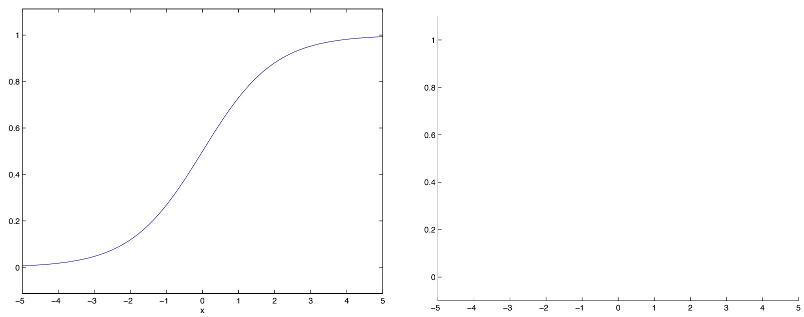


We can use this network to classify our inputs by assigning any input for which the output is greater than or equal to 1/2 to class 1 and for which the output is less than 1/2 to class 2. Using a weight vector of [0 1 1], compute the class for each of these input vectors:

input	class
[-1 -1]	
[-1 1]	
[1 -1]	
[1 1]	

2 Separator

1. Notice that in the last problem, we successfully created a separator for XOR, which we were unable to do with the sigmoidal transfer function. In this section we will see why this is the case. On the left, below, is a plot of the sigmoid as a function of its input. On the right, draw the corresponding plot for the Gaussian transfer function we used above.



2. Which of the 3D plots in Figure 1 corresponds to this Gaussian function as a function of a two-dimensional input?

3. So how can we describe the separator that we get with a simple 1-layer network using this transfer function? Can we describe it with (an) equation(s)?

3 Different Separators

Figure 2 shows three separators for the same set of data points. Each attempted separator uses a different kernel function.

(a) What function is being plotted in Figure 2? Where is it positive? Negative? What is the decision boundary in each diagram? Where are the support vectors?

(b) Fill in the table below, indicating which kernel functions might have been used:

Kernel	Attempted Separator
Linear	
Polynomial (n=2)	
Radial Basis Function	

4 Post-Training Calculations

Assume that we are using an SVM with a **polynomial kernel of degree 2**, i.e. $k(x^i, x^j) = (1 + x^i \cdot x^j)^2$. Say that training produces the following support vectors for each of which the α value is equal to 0.05.

x_1	x_2	y
-1	2	+1
1	2	-1

(a) What is the value of b?

(b) What value does this SVM compute for the input point $(1, 3)$?

5 Hand Training

What are the values for the α_i and the offset b that would give the maximal margin linear classifier for the two data points shown below. You should be able to find the answer without deriving it from the dual Lagrangian.

i	x^i	y^i
1	0	+1
2	4	-1

6 RBF Complexity

Consider the one-dimensional classification problem defined by the following data points. Imagine attacking this problem with an SVM using a radial basis function kernel.

i	x^i	y^i
1	1	+1
2	2	-1
3	3	-1
4	4	+1
5	5	+1

(a) Which of the above data points will be support vectors?

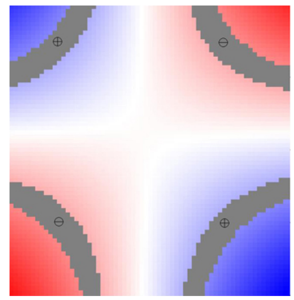
(b) Assume that we want the classifier to return a positive output for the +1 points and a negative output for the -1 points. Draw a plausible classifier output for every feature value in the interval $[0, 6]$. Do this twice. Once, assuming that the standard deviation σ is very small relative to the distance between adjacent training points. And again, assuming that the standard deviation σ is approximately the distance between adjacent training points. Note that a Gaussian kernel is close to zero for values farther than three standard deviations from its center.

(c) Can you conclude anything from these sketches about the complexity of the RBF kernel as a function of its standard deviation σ?

7 XOR Using an RBF

Given the following four data points, we find a separator using a radial basis function (RBF) with $\sigma = 1$ that produces the following plot:

i	x^i	y^i
1	[-1 -1]	-1
2	[-1 +1]	+1
3	[+1 -1]	+1
4	[+1 +1]	-1



(a) Recall that the equation that determines the distance from the separator is:

$$h'(u) = \sum_{i=1}^{k} \alpha_i y^i K(x^i, u) + b$$

$$K(x^i, u) = e^{-(||x^i - u||)^2/(2\sigma^2)}$$

t	e^t
-1	0.3679
-5	0.0067
-9	0.0001

The α values are all 1.33 in this example and the offset is 0. Assume you are given a **test** point [+2 +2]. What is the distance from the separator (use the e^t values from the table above)?

(b) Consider what we needed to do to compute this separator. We first needed to optimize the α values. To do this, we needed to compute $K(x^i, x^j), \forall i, j$. Compute the kernel outputs using the same radial basis function ($\sigma = 1$) for the points below.

K(i, j)	kernel value
K(1, 1)	
K(1, 2)	
K(1, 3)	
K(1, 4)	

(c) Sketch a picture of how the plot would change if we had been given a fifth **training** point $(y^5, x^5) = (-1, [+2\ +2])$? What if the fifth training point had been $(y^5, x^5) = (-1, [+\frac{1}{2}\ +\frac{1}{2}])$ instead?

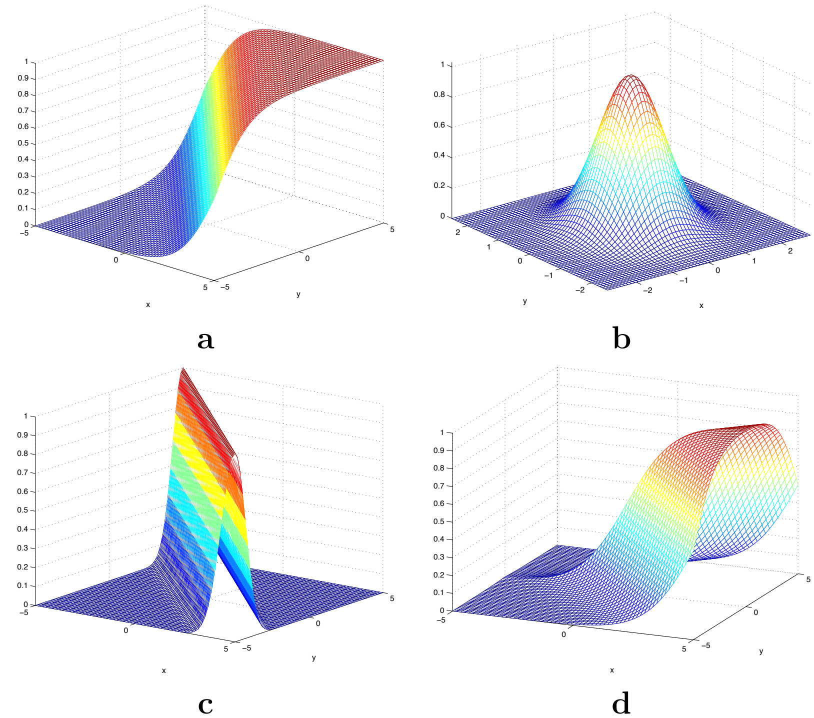


Figure 1: Problem 2 (Separator)

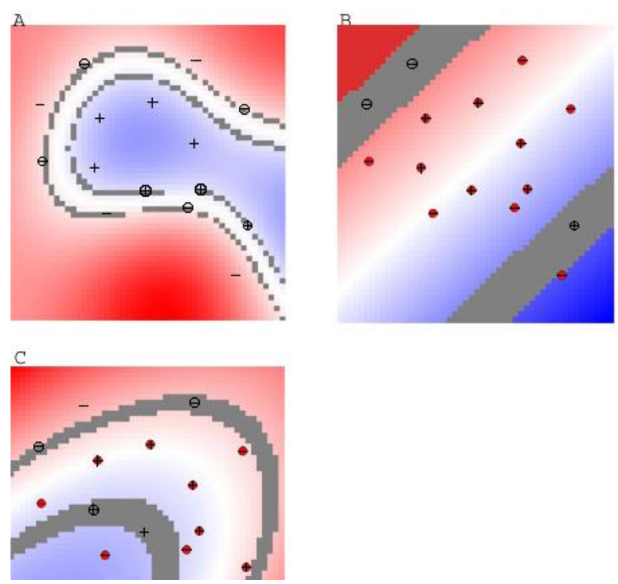


Figure 2: Problem 3 (Different Separators)