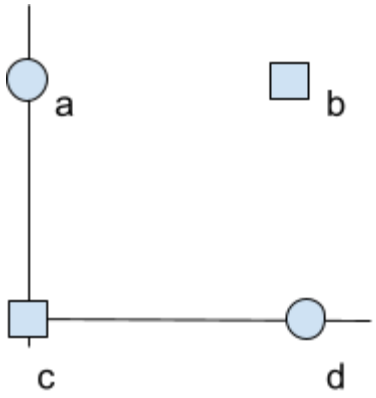
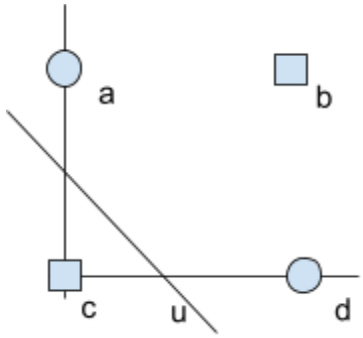
Consider the data below (also known as the XOR function). Clearly no single line separates the squares from circles. In order to classify the data below we will create a new feature space from simple linear classifiers.



Consider the linear classifier u below.

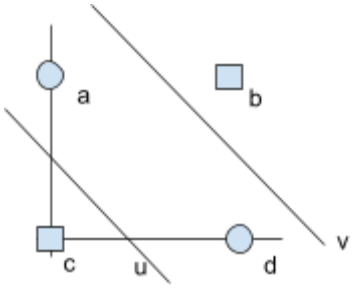


The (inactivated) output of each point above is given by $u^T x$. We need to non-linearize this output in order to get non-linear classification. If we just use the output as above then our final classifier is simply another linear classifier, and that will fail on XOR. In our example here we will use the sign(x) function activation.

The advantage of using sign(x) is easy to illustrate examples and workout examples by hand. But sign(x)=+1 (if x>0 and -1 if x<=0) is non-differentiable and therefore we cannot do gradient descent with it. Thus in practice we use other activation functions such as sigmoid(x)= $1/(1 + e^{-x})$ and relu(x)=max(0,x) both of which are differentiable. In the case of relu we use the sub-gradient.

According to u the outputs of a, b, c, and d are

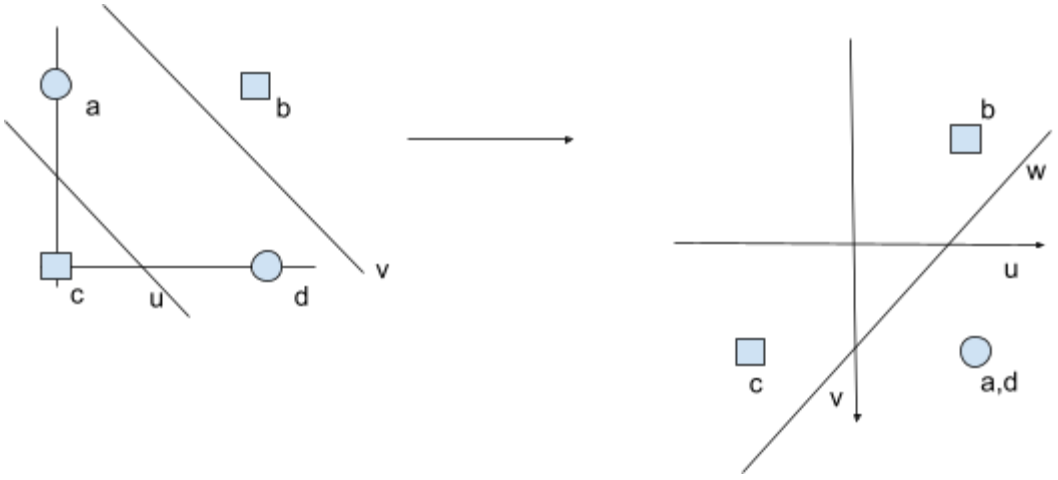
	u
a	1
b	1
c	-1
d	1



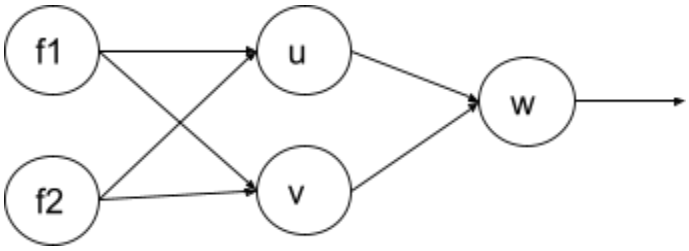
According to v the outputs of a, b, c, and d are

	u	v
a	1	-1
b	1	1
c	-1	-1
d	1	-1

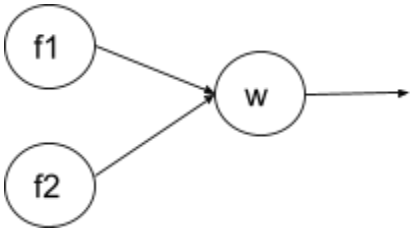
We interpret the above output as coordinates of our points in a new space given by axes u and v. We can see that in the new space it's easy to classify our data with a simple linear classifier w.



We can represent the above procedure in a neural network diagram shown below:



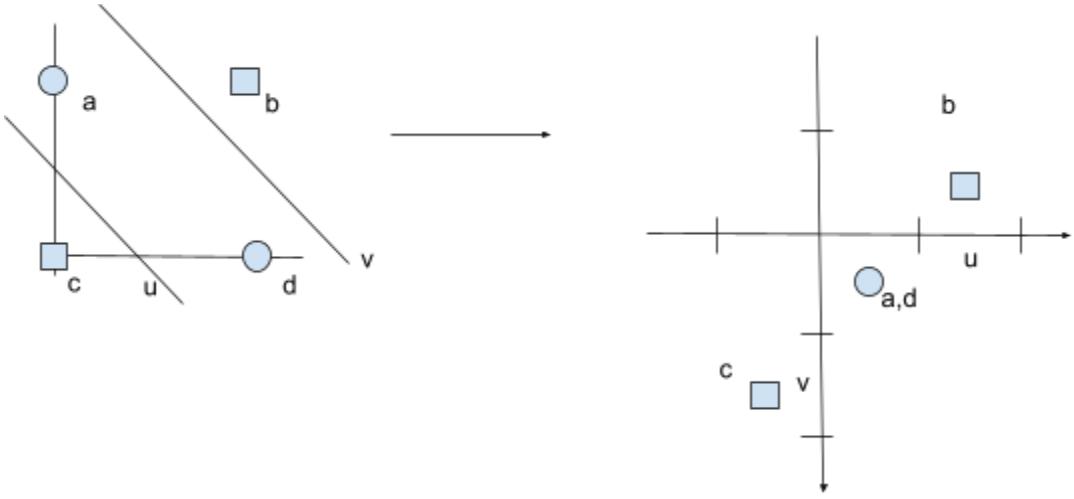
If there was no hidden layer and just one hyperplane w in the original feature space then we would describe that with the network below:



In the above example we have u=(1,1), u_0=-.5, v=(1,1), v_0=-1.5. Let us see the outputs of u and v without taking the sign.

x	y	u^Tx+u_0	v^Tx+v_0	sign(u^Tx+u_0)	sign(v^Tx+v_0)
a=(0,1)	1	.5	-.5	1	-1
b=(1,1)	-1	1.5	.5	1	1
c=(0,0)	-1	-.5	-1.5	-1	-1
d=(1,0)	1	.5	-.5	1	-1

Without activating the outputs our data in the new feature space looks like below. All four points are on one line now and cannot be separated in the new space.



The above example is to illustrate the importance of activating the output.