Multiclass classification is an extension of binary. There are three main multiclass strategies in machine learning:

1. One vs all:
2. One vs one
3. Hierarchical tree-like determination of classes

Suppose we have data from four classes:

	Y	f1	f2
X0	0	1	2
X1	0	2	2
X2	1	20	21
X3	1	22	19
X4	2	5	6
X5	2	7	5
X6	3	-1	-2
X7	3	-2	-3

<u>One vs all</u>

We create four new datasets:

D0:
x0, x1 have label +1
x2, x3, x4, x5, x6, x7 have label -1

D1:
x2, x3 have label +1
x0, x1, x4, x5, x6, x7 have label -1

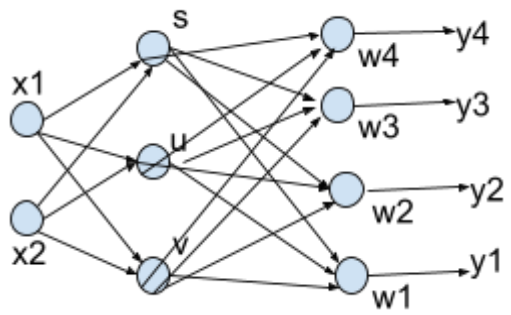
D2:
x4, x5 have label +1
x0, x1, x2, x3, x6, x7 have label -1

D3:
x6, x7 have label +1
x0, x1, x2, x3, x4, x5 have label -1

<u>Training</u>: We learn four binary classifiers on each dataset D0, D1, D2, and D3 giving us C0, C1, C3, and C4.
<u>Testing</u>: We assign a test datapoint x' to the class i with the largest output value of Ci(x'). In the above example we calculate C0(x'), C1(x'), C2(x'), and C3(x'). We then assign x' to the class with the highest output value.

How do we perform multiclass in a neural network? We extend one-vs-all to the final layer.

Below is a neural network with four nodes in the final layer. Each node corresponds to a one-vs-all binary classifier.



In training the above network on the data below we would set the y labels to be one-hot label vectors.

	Y	f1	f2
X0	(1,0,0,0)	1	2
X1	(1,0,0,0)	2	2
X2	(0,1,0,0)	20	21
X3	(0,1,0,0)	22	19
X4	(0,0,1,0)	5	6
X5	(0,0,1,0)	7	5
X6	(0,0,0,1)	-1	-2
X7	(0,0,0,1)	-2	-3

The output of a test datapoint x’ from the above network would be a vector containing four values (y1, y2, y3, y4). We select the class with the highest value.