Q1

If there were no bias weights, all linear activation functions (represented as straight lines in the coordinate system) would have to pass through the origin. By adding bias weights, the straight lines would be able to move freely in the coordinate system, making it easier to delineate the data domain.

Q2

Discriminant function: Figure 1 does not distinguish the distribution of data well. Figure 2 and Figure 3 successfully distinguish the distribution of data, and Figure 2 is more accurate than Figure 3.

Training and test errors: Figure 1 has the worst training and test errors among the three Figures and Figure 2 and Figure 3 are much better than Figure 1, Figure 3 is slightly better on training error than Figure 2 but slightly worse on test error.

Generalisation performance: Figure 1 has the worst generalisation performance among the three Figures, Figure 2 and Figure 3 have better generalisation performance, Figure 2 is more accurate than Figure 3.

Q4

The activation functions used in Figure 1-3 are linear functions, so the discriminant functions represent polygons, the activation function used in Figure 4 is curvy function, so the discriminant function represents curved boundaries.

Q5

op1 = -6:0.1:6

op2 = 0.2

op3 = 0.2

y = zeros(1, size(op1, 2))

for n = 1 : size(op1, 2)

y(n) = exp(op1(n)) / (sum(exp(op1(n)) + exp(op2) + exp(op3)))

end

plot(op1,y)

xlabel("op1")

ylabel("y")

