Assignment: A.2 Classifier NNs Name: Alisher Siddikov Class: MSDS 458
Date: 02/16/2020

Introduction

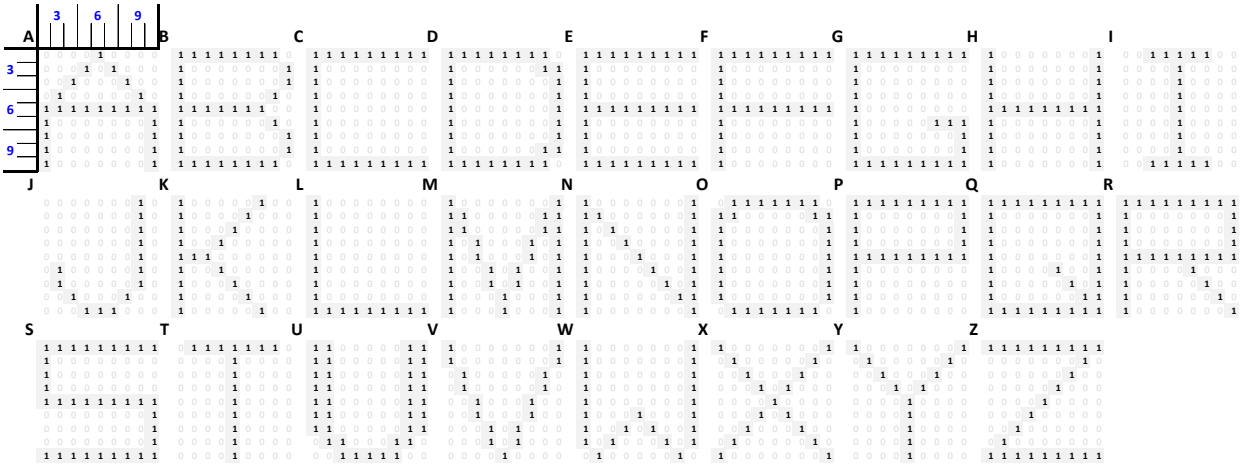
This assignment is similar to the previous assignment. We are asked to explore how the hidden layers in a simple single-hidden layer network learn to represent features within the input data. Input data is an array form of 26 capital letters of the English alphabet. Each letter has 9x9 pixels, and this is transformed into a vector of length 81 giving one input node for each pixel in the letter. You can see 9x9 pixeled capital letters in Figure 1.

Therefore, by default, we have the following parameters:

- random.seed(12345)
- alpha = 1.0
- eta = 0.5
- maxNumIterations = 5000
- epsilon = 0.01
- noise (all) = 0.00
- numHiddenNodes = 6

During the analysis (without noises), we will see that hidden nodes cluster letters by identifying learned features such as finding mutual patterns: square, vertical, horizontal, rounded corners, diagonals, and middle lines. We will identify which hidden nodes responded to which classes – and identify which hidden nodes responded to classes that had similar letter shapes. We will train our model to put our 26 pixelated letters into the correct class assignments. For example, letters 'D', 'O' and 'Q' can be grouped to 'O' like shape because they all seem to share similar pixels.

Analysis before tweaking the pixels and classes



Assignment: A.2 Classifier NNs Name: Alisher Siddikov Class: MSDS 458
Date: 02/16/2020

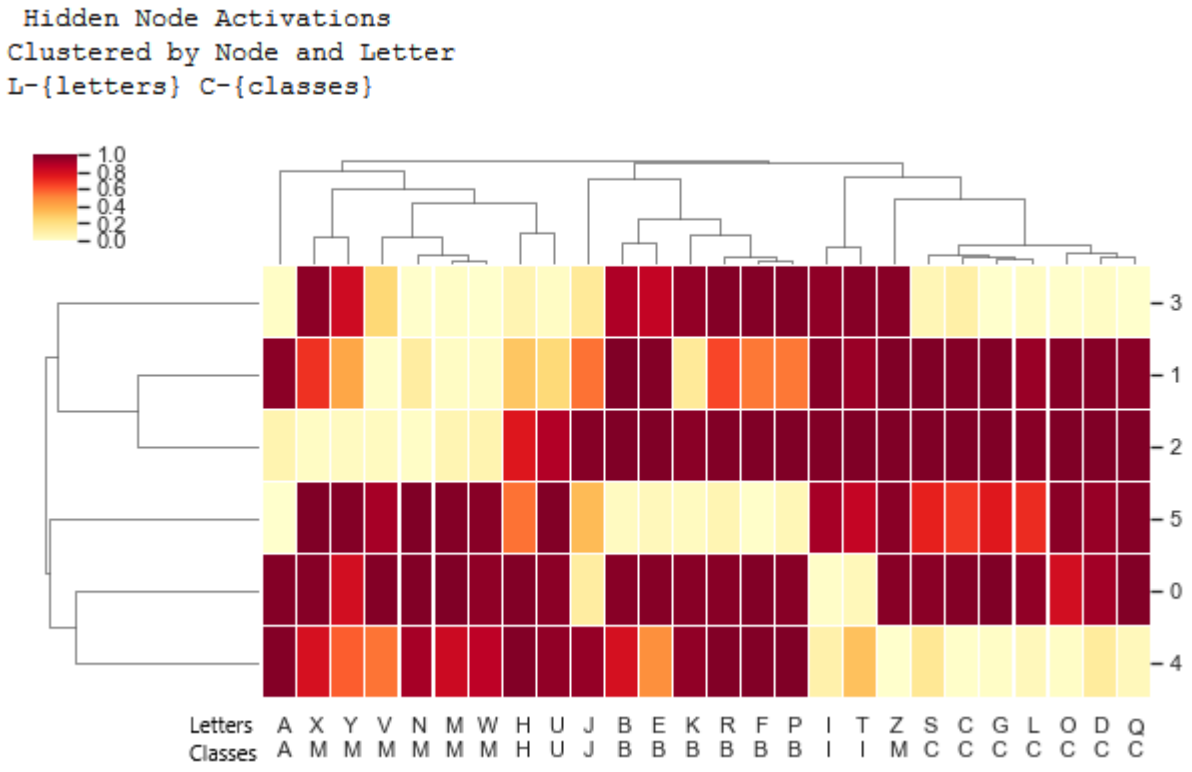
By default, there are seven classes. As you can see in the accuracy table below, the letters A, H, J, and U have their own classes and letter S is under the C class. We should reclassify them and re-run the code again.

The accuracy is 92.31: H and U letters are misclassified.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Letter	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Desired	0	1	2	2	1	1	2	3	4	5	1	2	6	6	2	1	2	1	2	4	7	6	6	6	6	6
Selected	0	1	2	2	1	1	2	6	4	5	1	2	6	6	2	1	2	1	2	4	6	6	6	6	6	6

Percent Accuracy 92.31

Here is the dendrogram heatmap. It suggests A and U letters should be under M class. And letters I, T, and Z should be under the C class. Before we reclassify them, we need to tweak the letter pixels.

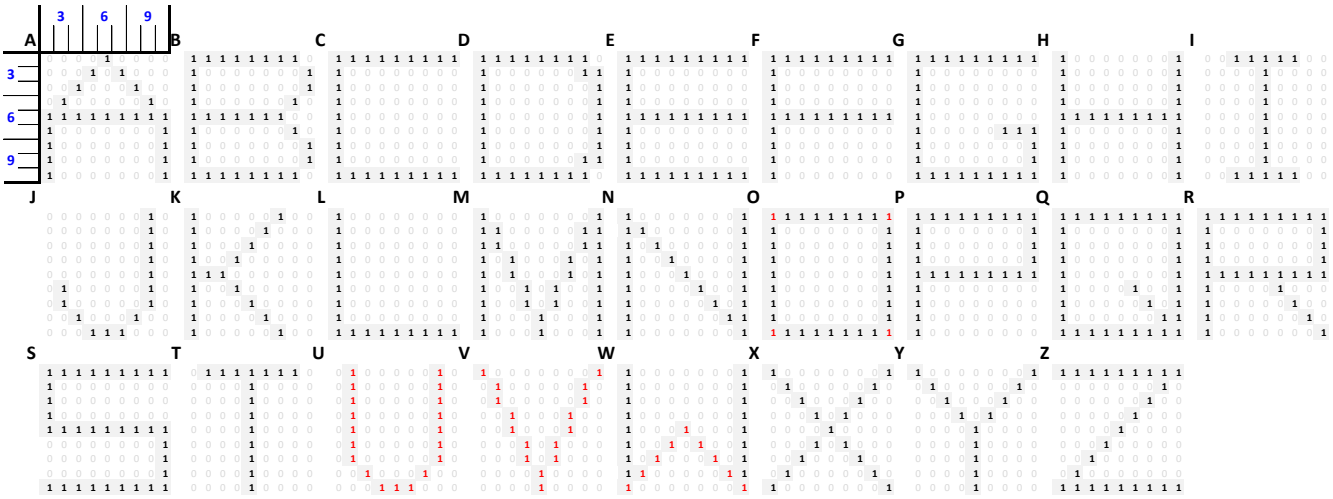


Hidden Node	Activated Strongest	Activated Weakest
0	all letters except letters J, T, I A-Z except for J, T, I	when the letters which have vertical lines in the middle or RHS only: J, T, I
1	when the letters contain horizontal lines at the top *and* the bottom: Z, S, B, G, E, C, D, I, O, Q	when the letters don't share horizontal lines at the top or bottom: Y, H, U, K, N, M, W, V
2	when the letters contain horizontal lines at the top, the middle, *or* the bottom; It is similar to hidden node 1, but it also includes the following	when the letters don't share horizontal lines at the top and the bottom and have some diagonals:

Assignment: A.2 Classifier NNs Name: Alisher Siddikov Class: MSDS 458
Date: 02/16/2020

Hidden Node	Activated Strongest	Activated Weakest
	P, F, R, L, T, K, J, U letters	H, W, A, M, V, Y, X, N
3	when the letters contain vertical line LHS and middle horizontal line: P, F, R, K, B, E and the letters which share vertical lines: X, Y, Z and middle vertical line: I, T	when the letters contain vertical line RHS: J, U, H, M, N, W, and C classes
4	when the letters contain horizontal lines at the middle and no bottom horizontal line: P, H, R, F, A, K and the letters contain corners such as U, J It also activated for the letters which include horizontal lines and diagonal such as N, W, M	when the letters share bottom horizontal line such as I and C classes
5	when the letters contain diagonals: N, X, M, Y, W, Z, V and two vertical lines: N, U, M, W, O, D and somewhat stronger activated when the letters contain a vertical line with a top or bottom horizontal line: I, T, G, L	when the letters share middle horizontal lines such as A, H, and B classes

Analysis after tweaking the pixels and classes – Classification # 1



We left the default parameters as it is. In this assignment, we created big shape classes for our letters, and we should make sure that they were classifying the way we think they should. We have letters that do not seem to belong in a class; we will try another class and/or create a new class.

We can see that O, U, V, and W letters should be standardized to group them correctly. I made some pixel corrections on letters O, U, V, and W, and the pixels are highlighted in red ink. Now O is similar to Q, U is similar to J, V is similar to M, and W is similar to X.

I also reclassified letters. I included letters A, H, and S to class 0 as well as J and U to class 2.

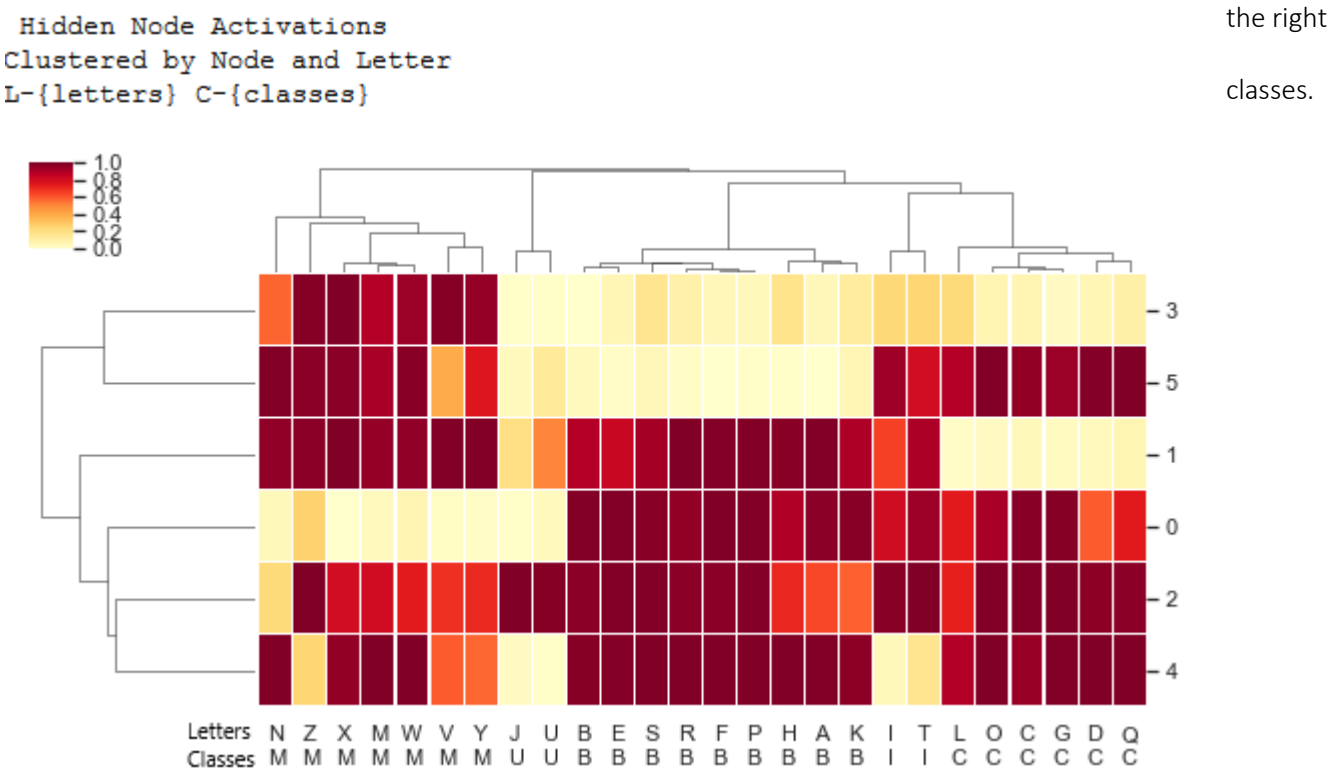
- Class 0 or B: (contains middle horizontal line): A, B, E, F, H, K, P, R, S
- Class 1 or C: (contains at least vertical line at LHS and horizontal line at the bottom, mostly squarish): C, D, G, L, O, Q
- Class 2 or U: (contains rounded corners): J, U
- Class 3 or M: (contains diagonals): M, N, V, W, X, Y, Z
- Class 4 or I: (contains middle vertical line): I, T

We achieved an accuracy score of 100 by fixing pixels and reclassifying the letters.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Letter	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Desired	0	0	1	1	0	0	1	0	4	2	0	1	3	3	1	0	1	0	0	4	2	3	3	3	3	3
Selected	0	0	1	1	0	0	1	0	4	2	0	1	3	3	1	0	1	0	0	4	2	3	3	3	3	3

Percent Accuracy 100.00

Here is the dendrogram heatmap of classification 1. The horizontal branches of the dendrogram cluster the letters into classes in a nice way as we expected. We cluster I and T letters with C group because they share horizontal lines at the top and bottom. We can see that our model generalized well enough to include letters in the right classes.



Assignment: A.2 Classifier NNs Name: Alisher Siddikov Class: MSDS 458
Date: 02/16/2020

Hidden Node	Activated Strongest	Activated Weakest
0	when the letters contain vertical line at LHS *and* have the middle horizontal line: F, P, E, B, G, K, S, A, R, H when the letters contain vertical line *and* horizontal line at the top *and/or* bottom: C, T, O, I, L, Q, D	when the letters don't share horizontal lines at the top *and* bottom and do not have a middle line (two or more middle pixels) such as M and U classes: W, N, U, M, V, Y, J, X
1	when the letters contain diagonals: X, R, Y, A, V, Z, N, W, M, K or middle horizontal line (four or more middle pixels): P, F, H, S, B, E or middle vertical line I, T	when the letters do not have diagonal or middle horizontal lines (four or more middle pixels) *but* have rounded corners or squares such as U and C classes U, J, Q, C, D, G, O, L
2	all letters except letters A, K, N A-Z except for A, K, N	when the letters do not have a bottom horizontal line, *but* have a negative diagonal line and have a vertical line at LHS: A, K, N
3	when the letters contain diagonal lines such as M class: X, Z, V, Y, W, M, N	when letters do not share diagonal lines such as I, C, B, and U classes
4	all letters except letters V, Y, Z as well as classes I, U A-Z except for V, Y, Z, T, I, J, U	when letters do not share a diagonal line with vertical line borders: V, Y, Z *or* do not share square as well as middle horizontal line such as I and U classes: T, I, J, U
5	when the letters contain outer walls, *or* diagonal, lines, *or* vertical middle line, *but* no middle horizontal line such as C, M, I classes: Q, O, D, G, C, L, N, M, W, X, Z, I, T	when letters contain middle horizontal lines and rounded corners such as B and U classes

Analysis after tweaking the pixels and classes – Classification # 2

We left almost all default parameters the same except the hidden and output nodes. We changed the hidden node from 5 to 8 and the output node from 4 to 9. We also reclassified letters into smaller classes. The reclassification was intended to get better identification of feature nodes and better output classification.

- Class 0 or H: (letters contain middle horizontal line but no horizontal outer walls): A, H, K
- Class 1 or F: (contains vertical line at the LHS and horizontal line at the top and middle): F, P, R
- Class 2 or B: (contains horizontal lines at the top, middle, and bottom): B, E, S
- Class 3 or C: (contains square but have some open walls): C, G, L
- Class 4 or O: (contains square and no open walls): D, O, Q
- Class 5 or U: (contains rounded corners): J, U
- Class 6 or M: (contains V shape in the middle): M, V
- Class 7 or N: (contains diagonal lines within two vertical walls): N, W

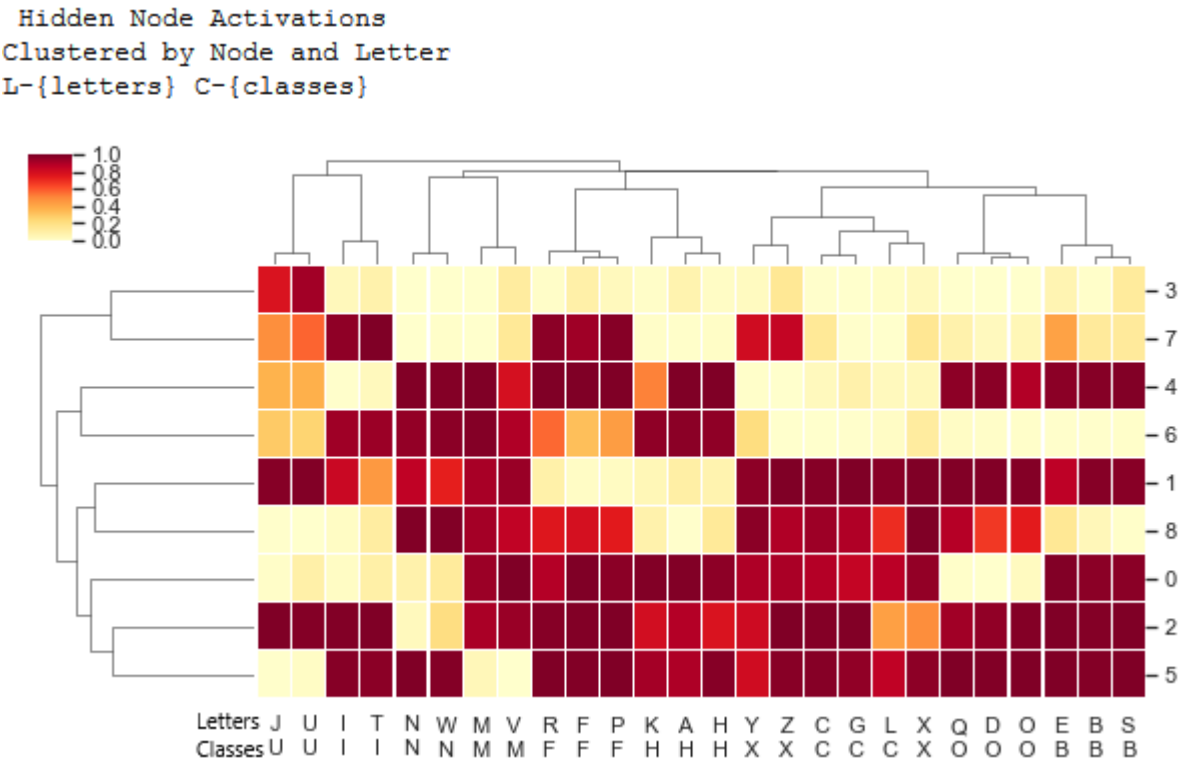
- **Class 8 or X:** (contains diagonal lines without two vertical walls): X, Y, Z
- **Class 9 or I**: (contains middle vertical line): I, T

The accuracy score did not change.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Letter	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Desired	0	2	3	4	2	1	3	0	9	5	0	3	6	7	4	1	4	1	2	9	5	6	7	3	3	3
Selected	0	2	3	4	2	1	3	0	9	5	0	3	6	7	4	1	4	1	2	9	5	6	7	3	3	3

Percent Accuracy 100.00

Here is the dendrogram heatmap of classification 2. Since we have smaller classes, X, Y, Z letter which classified as X class is not optimal by looking at dendrogram. Next steps if I had a time at this point, I need to take a look at why the letter X grouped with the letter L. I was expecting the letters X and Y are closer to each other than letters X and L.



Hidden Node	Activated Strongest	Activated Weakest
0	all classes except I, N, O, U classes and here are explanations of each class. the strongest activation found in B, F, H, M classes. B, F, H classes have a horizontal line at the middle: E, S, B, F, P, R, A, K, H The second strongest activation found in X and C classes. X class contains diagonal lines without two vertical walls such as	when the letters contain a middle vertical line such as I, T when the letters don't share the middle horizontal line but have two (or 1.5 for J) vertical walls: W, N, O, Q, D, U, J

Assignment: A.2 Classifier NNs Name: Alisher Siddikov Class: MSDS 458
Date: 02/16/2020

Hidden Node	Activated Strongest	Activated Weakest
	X, Y, Z C class contains square but have some open walls such as C, G, L	
1	all classes except for H, F classes	when letters contain the middle horizontal line but do not have a bottom horizontal line: A, R, H, K, P, F
2	all classes except N class as well as X, L letters the second strongest activation found mainly in H class or letters which do not fit other classes: A, K, H, Y	when letters contain diagonal lines within two vertical walls or without two horizontal walls: W, N
3	it activates only U class which letters contain rounded corners: J, U	all classes except for U class
4	all classes except for X, C, I, U classes	when letters contain diagonal lines without two vertical walls such as X class: X, Y, Z or when letters contain square but have some open walls such as C class: C, G, L or when letters contain a middle vertical line such as I class: I, T or somewhat weaker when letters contain rounded corners such as U class: J, U
5	all classes except for M, U classes	when letters contain V shape in the middle such as M class: M, V or when letters contain rounded corners such as U class: J, U
6	when letters contain V shape in the middle such as M class: M, V or when letters contain diagonal lines within two vertical walls or without two horizontal walls: W, N or when letters contain middle horizontal line but no horizontal outer walls; do not fit other classes such as H class: A, H, K or when letters contain a middle vertical line such as I class: I, T the second strongest activation is found in F class: P, R, F	when letters contain diagonal lines without two vertical walls such as X class: X, Y, Z or when letters carry the vertical line at the LHS and horizontal line at the bottom such as O, C, B classes: D, O, Q, L, G, C, S, B, E

Hidden Node	Activated Strongest	Activated Weakest
7	when letters contain vertical line and horizontal line at the top but no horizontal full line at the bottom such as I and F classes: F, P, R, I, T the second strongest activation is found when letters have a half diagonal line or rounded corners such as X and U classes: Z, Y, U, J	all classes except I, F, X, U classes
8	when letters contain diagonal lines or at least ¾ squares: X, W, N, Y, M, Z, V when letters contain square shapes: C, G, Q, O, L, D or when letters contain full vertical line at the LHS, and horizontal lines at the top and middle such as F class: F, P, R	when letters have middle line excluding F class: E, H, K, B, S, A when letters contain a middle vertical line or rounded corners: T, I, J, U

Conclusion

In this assignment, we created big shape classes for our letters, and we made sure that they were classified the way we thought they should. We had letters that did not seem to belong in a class, so we reclassified them. We identified which hidden nodes responded to which classes – and identified which hidden nodes responded to classes that had specific pixels (of letter shapes) in common. We trained our model to put our 26 pixelated letters into the correct class assignments. Classification # 1 hidden node activations made sense, and the letters aligned to appropriate classes in the dendrogram heatmap. Classification # 2 hidden node activations made sense; the reclassification was intended to get better identification of feature nodes and better output classification. However, there appeared to be one of the classifications such as the letter X failed to identify features as it was done classification # 1. I provided an explanation of the features the hidden nodes were activating on. The increasing number of classes and the number of hidden nodes did not result in a better model.