CS188–Winter 2020 — Homework 2 Solutions

Eugene Choi, SID 905368197

Collaborators: Luca Matsumoto (204794180) , Trevor Holt (204794180)

1.Perceptron Training

								w0	w1	w2	w3
							initial weights	1	1	1	1
x0	x1	x2	x3	t	z (predicted label)	delta w					
1	1	0	1	0	1	-1		0	0	1	0
1	1	1	0	0	1	-1		-1	-1	0	0
1	1	0	1	1	0	1		0	0	0	1
1	0	1	1	1	1	0		0	0	0	1
1	1	0	1	0	1	-1		-1	-1	0	0
1	1	1	0	0	0	0		-1	-1	0	0
1	1	0	1	1	0	1		0	0	0	1
1	0	1	1	1	1	0		0	0	0	1
1	1	0	1	0	1	-1		-1	-1	0	0
1	1	1	0	0	0	0		-1	-1	0	0
1	1	0	1	1	0	1		0	0	0	1
1	0	1	1	1	1	0		0	0	0	1

(a) After several iterations of the training set, I found that the weights never converged. The reason for inseparability is because the input is not linearly separable, therefore, perceptron does not end up converging.

2.Input Validation

(a) Approach 1. Plotting your stream of data: After plotting your stream of data, you can check to see if there is a certain type of trend (a trend being a type of pattern that your data creates that can be fit to the graph of a certain function – examples being linear, logarithmic, or exponential) that the data follows and see if there are many outliers with your data when finding the trend. If your data does not follow a trend and you find many outliers, then there is a high chance that the data obtained is invalid.

Approach 2. Create a threshold that your data should not exceed or fall under: Using historical data, you can determine a threshold that your data should fit into to see if the data that the SickBit sensor is obtaining is valid. For example, if you are taking the body temperature of a living human beings and you obtain values such as 30 degrees Celsius and 50 degrees Celsius, then your stream of data is most likely invalid because the human body temperature should always lie between these temperatures:
1) 36.5–37.5 °C (97.7–99.5 °F) for normal,
2) $>$ 37.5 or 38.3 °C (99.5 or 100.9 °F) for fever,
3) $<$ 35.0 °C (95.0 °F) for hypothermia.

3.Distributions

(a) i)
(bar width * Percent per inch for Mother's height) + (bar width * Percent per Inch for Mother's height)
(2 Inches * 6 Percent per Inch) + (2 Inches * 14 Percent per Inch) = 40 Percent

ii)
Unknown because we are not told how each height is partitioned within the bin. For example, within the bin from 66 to 68, there might be no fathers with a height of 67 inches, but instead only fathers with a height of 68 inches.

iii)
Unknown because we do not know the number of sons that exist within a family. Although there is only one mother and father per family, the number of sons can be greater than one. Since this is the case, we cannot accurately determine the size of any subset for the son.

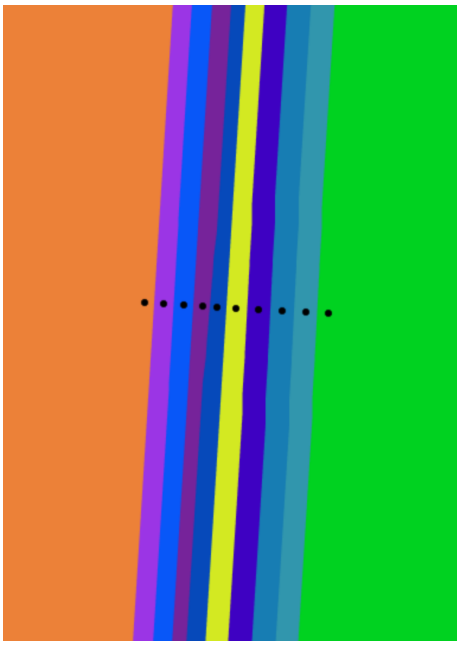
iv)
100 Percent - (2 Inches * 2 Percent per Inch) = 96 Percent
.96 * 200 Mothers = 192 Mothers

(b) 2 Inches *(6 Percent per Inch + 2 Percent per Inch) = 16 Percent
Height of the bar would be = 16 Percent / 4 Inches = 4 Percent per Inch (the four inches comes from the combined width of the two previous bins.

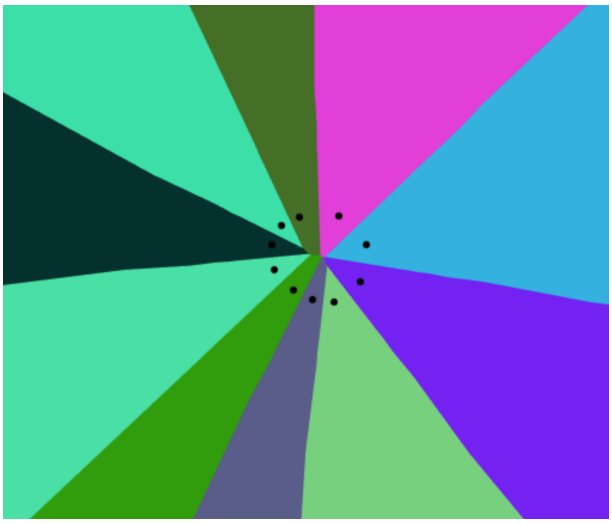
(c) By looking at the histogram, we can see that 70 to 72 inches is the maximum height that a mother can be. The percentage of sons above 70 inches is (2 Percent per Inch + 8 Percent per Inch + 14 Percent per Inch) * 2 Inches = 48 Percent. The percentage of sons above 72 inches is (8 Percent per Inch + 2 Percent per Inch) * 2 Inches = 20 Percent. Therefore, the percentage of sons that are taller than all of the mothers is between 20 Percent and 48 Percent.

4. Voronoi

(a) Voronoi diagram of 10 points all on a line



Voronoi diagram of 10 points all on a circle



Summary: What these two diagrams have in common are that they both have ten points that are used to partition a plane into regions. The regions come from splitting the plane directly down the half way distance from one point to the next.

5.Augmentation

(a) Two features that would be interesting to cross from Project 2 are age and cp (chest pain type). I feel that these features would be interesting to cross because we can see if the assumption of chest pains becoming a more common occurrence in individuals of older age is valid.

(b) To feature cross latitude and longitude, you would have to bin them because they are continuously valued. By binning them, you would be able to categorize the values into ranges and cross the ranges rather than than trying cross with individual values.

(c)

x	y	z
0	0	1
1	1	1
2	2	1
0	3	0
1	4	0
1	0	0
2	1	0

(d) Whenever I tried to print the table as the answer for (c), it would not let me unless I created this extra item (d).