CS 188 HW 2

1. **Instructions:**
    a. **Study Partners:**
        i. None

2. Perceptron Training

Note that this train set is not linearly separable, which can be seen by the first and third rows having a different output for the same input. As such, I've shown the weights after each pattern in Table 1 on the next page, but the weights of the perceptron won't converge (though the accuracy will).

x_0	x_1	x_2	x_3	t	w_0	w_1	w_2	w_3	Net	Output	Δw_0	Δw_1	Δw_2	Δw_3
1	1	0	1	0	1	1	1	1	3	1	-1	-1	0	-1
1	1	1	0	0	0	0	1	0	1	1	-1	-1	-1	0
1	1	0	1	1	-1	-1	0	0	-2	0	1	1	0	1
1	0	1	1	1	0	0	0	1	1	1	0	0	0	0

3. **Input Validation**

- Range limits:
  - We can make use of prior knowledge or histograms of historical data in order to set upper limits and lower limits for what is considered a valid reading. This is an effective approach for blood pressure, heart rate, body temperature, and other vital signs for which there are known limits.
- Statistical tests:
  - You can use statistical methods to compare a set of readings against a historical dataset. For example, you can compare the variance or the difference between 25th and 75th percentiles for recent data against historical data in order to determine if the recent data is valid.

4. **Distributions**

a.

i. $2 * 6\% + 2 * 14\% = 40\%$

ii. Unknown

iii. $(2 * 14\% + 2 * 8\% + 2 * 2\%) = 48\%$ --> Unknown (don't know total # of sons)

iv. $(2 * 6\% + 2 * 14\% + 2 * 15\% + 2 * 10\% + 2 * 2\% + 2 * 1\%) * 200 = 192$

b. $\frac{2*6\%+2*2\%}{4} = 4\%$

c. The percentage of sons that are taller than all of the mothers is between 20 and 48.

We know all mothers are less than 72 inches so at least the $2 * 8\% + 2 * 2\% = 20\%$ of sons taller than 72 inches are taller than all mothers. For the 70-to-72 bin, if all $2 * 14\% = 28\%$ of sons in that bin are taller than all mothers in that bin, then $2 * 8\% + 2 * 2\% + 2 * 14\% = 48\%$ of sons would be taller than all mothers.

5. **Voronoi**

Diagram 1 - Voronoi diagram of 10 points all on a line:

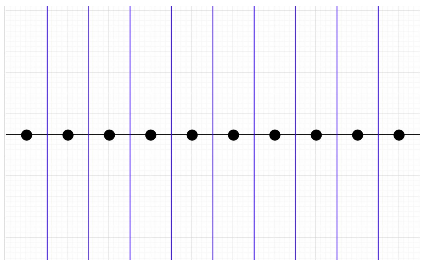
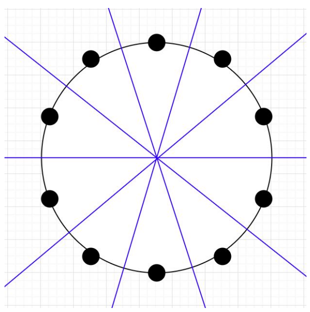


Diagram 2 - Voronoi diagram of 10 points all on a circle:



These two diagrams have in common that all Voronoi cells/regions extend to the boundary of the defined plane. Also, aside from the center point in Diagram 2, both diagrams contain only line segments (points in the plane that are equidistant to the two nearest seed points) instead of Voronoi vertices (points equidistant to three or more seed points). Also, the line between each pair of points is a perpendicular bisector, and there are n-1 lines separating n points.

6. Augmentation

a. It might be interesting to cross sex and age from Project 2 together, because in my analysis they were only moderately positively correlated to being sick with heart disease. However, if specific gender and age (or age group) combinations are especially susceptible/strongly correlated to heart disease, this would allow the model to account for that.

b. In practice, machine learning models seldom cross continuous features. However, they do frequently cross one-hot feature vectors. We could bin latitude and longitude, producing separate one-hot 5-element feature vectors. So a given latitude and longitude could be represented as binned_lat = [0, 1, 0, 0, 0] and binned_long = [0, 0, 0, 1, 0]. A feature cross of these two feature vectors, binned_lat X binned_long would be a 25-element one-hot vector (24 zeros and 1 one). The single 1 in the cross represents a particular conjunction of latitude and longitude.

c. Let class A represent sick kids and class B represent healthy kids. You can't draw a single straight line that neatly separates the sick kids from the healthy kids, so this is a nonlinear problem. Any line you draw will be a poor predictor of tree health, without feature crosses such as W = XY.

X	**Y**	**Z**
-4	4	A
-4	3	A
4	-4	A
4	4	B
4	5	B
-4	-4	B
-5	-4	B