Raymond Lin
304937942

CS 188 Data Science - Homework 2

1. Instructions:
   - partners: Derek Hu

2. Perceptron Training

Iteration	Prediction	Output
1	① $(1\cdot1)+(1\cdot0)+(1\cdot1)+1>0$	$\Rightarrow 1$ - incorrect
	$\Delta w_1 = (0-1)1$, $w_1 = 0$	
	$\Delta w_2 = (0-1)0$, $w_2 = 1$	
	$\Delta w_3 = (0-1)1$, $w_3 = 0$	
	$\Delta b = (0-1)1$, $b = 0$	
	② $(0\cdot1)+(1\cdot1)+(0\cdot0)+0>1$	$\Rightarrow 1$ - incorrect
	$\Delta b = (0-1)1$, $b=-1$	
	$\Delta w_1 = (0-1)1$, $w_1 = -1$	
	$\Delta w_2 = (0-1)1$, $w_2 = 0$	
	$\Delta w_3 = (0-1)0$, $w_3 = 0$	
	③ $(-1\cdot1)+(0\cdot0)+(0\cdot1)-1=-2$	$\Rightarrow 0$ - incorrect
	$\Delta b = (1-0)1$, $b=0$	
	$\Delta w_1 = (1-0)1$, $w_1 = 0$	
	$\Delta w_2 = (1-0)0$, $w_2 = 0$	
	$\Delta w_3 = (1-0)1$, $w_3 = 1$	
	④ $(0\cdot0)+(0\cdot1)+(1\cdot1)+0=1$	$\Rightarrow 1$ - correct
	- don't change weights	
2	① $(0\cdot1)+(0\cdot0)+(1\cdot1)+0=1$	$\Rightarrow 1$ - incorrect
	$\Delta b = (0-1)1$, $b=-1$	
	$\Delta w_1 = (0-1)1$, $w_1 = -1$	
	$\Delta w_2 = (0-1)0$, $w_2 = 0$	
	$\Delta w_3 = (0-1)1$, $w_3 = 0$	
	② $(-1\cdot1)+(0\cdot1)+(0\cdot0)+-1=-2$	$\Rightarrow 0$ - correct
	- don't change	

③ $(-1 \cdot 1) + (0 \cdot 0) + (0 \cdot 1) - 1 = -2 \Rightarrow 0$ - incorrect

$\Delta b = (1-0)1$, $b = 0$

$\Delta w_1 = (1-0)1$, $w_1 = 0$

$\Delta w_2 = (1-0)0$, $w_2 = 0$

$\Delta w_3 = (1-0)1$, $w_3 = 1$

④ $(0 \cdot 0) + (0 \cdot 1) + (1 \cdot 1) + 0 = 1 \Rightarrow 1$ - correct

- we can see that the weights are the same after each iteration
- at this rate, the perceptron algorithm will never converge
- thus, this dataset is not linearly separable

2. (cont.)

Iteration	x1	x2	x3	b/w0	w1	w2	w3
0				1	1	1	1
1	1	0	1	0	0	1	0
	1	1	0	-1	-1	0	0
	1	0	1	0	0	0	1
	0	1	1	0	0	0	1
2	1	0	1	-1	-1	0	0
	1	1	0	-1	-1	0	0
	1	0	1	0	0	0	1
	0	1	1	0	0	0	1

3. Input Validation

One technique we could use to check the validity of data is Schema Based Example Validation. This method identifies any anomalies in the input data by comparing statistics against an existing schema. The schema specifies what rules the data should follow, such as data type and categorical features.

Another technique is called Training-Serving Skew detection. In machine learning models that are deployed on applications, there exists a serving and training dataset. The serving data set is the one that contains data that is received from the day of service. The training data set is the one that is used to train a new model so that it remains updated, and it draws data from the serving data set. Since one data set depends on the other, Training-Serving Skew detection measures any inconsistencies between the two datasets.

4. a) i)

height range (in)	percent per inch	total percentage
60-62	6	12
62-64	14	28
		40

$\boxed{\text{Percentage of mothers that are at least 60 inches but less than 64 inches tall} = 40\%}$

ii)

height range (in)	percent per inch	total percentage
64-66	4	8
66-67	7	0-14
		8-22

$\boxed{\text{The percentage of fathers who have a height b/w 64-67 cannot be determined because we do not know the distribution of heights in the 66-68 bin}}$

iii)

height range (in)	percent per inch	total percentage
70-72	14	28
72-74	8	16
74-76	2	4
		48

$\boxed{\text{The number of sons can't be determined because we don't know the total number of sons in the entire dataset}}$

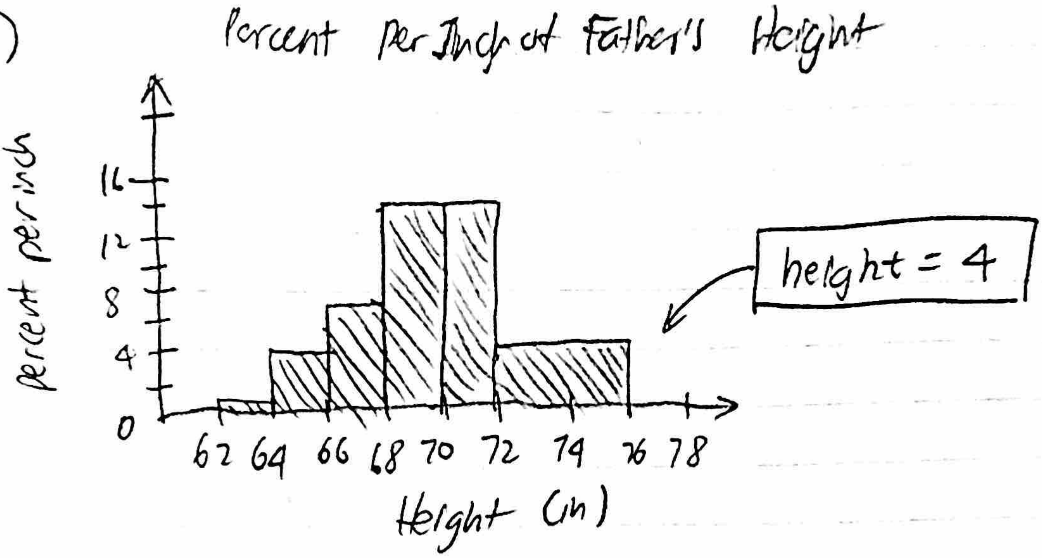
iv)

height range	percent per inch	total percentage
60-62	6	12
62-64	14	28
64-66	15	30
66-68	10	20
68-70	2	4
70-72	1	2
		96

$0.96 \times 200 = 192$

$\boxed{\text{Number of mothers at least 60in. tall} = 192}$

b)



c) Look at the 70-72 bin in both graphs: There are 2 possible extreme scenarios.

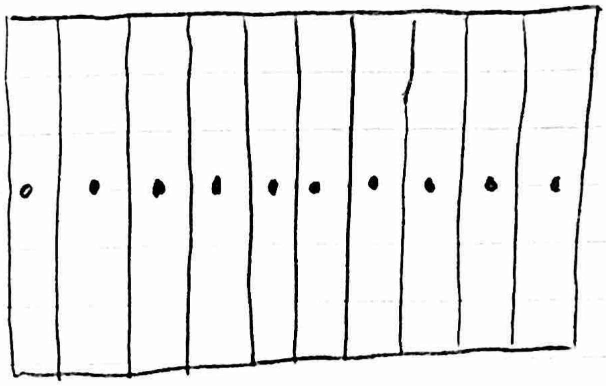
① If all mothers in that bin are shorter than all sons in that bin.
- percentage of sons taller than all mothers =
28 + 16 + 4 = 48%

② If all mothers in that bin are taller than all sons in that bin
- percentage of sons taller than all mothers
= 16 + 4 = 20%

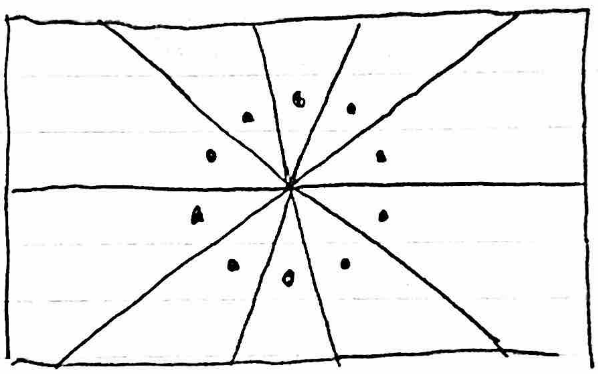
Thus, the percentage of sons that are taller than all the mothers is between 20% and 48%

5- Voronoi

- all 10 points on a line



- all 10 points on a circle



The two diagrams are split into 10 different regions of space.

6.

a)

It may be interesting to cross age and maximum heart rate achieved. Feature crossing is typically used in order to discover the joint effects of two different features, especially if a single feature doesn't have a large impact on prediction. Age by itself is probably not very indicative of heart disease, since many old people are healthy. Likewise, maximum heart rate may not be indicative of heart disease by itself. However, it would be interesting to see if both of those values together have a predictive impact. For example, if we want to check if people who are old and have low maximum heart rates are more likely to have disease, we can invert the maximum heart rates and then multiply that value with age. Thus, those people will achieve a very high score, while people with only one of those traits (either only old or only have low maximum heart rates) have a lower score.

It may also be interesting to see effects of crossing depression induced by exercise relative to rest and slope of the peak exercise ST segment. A depressed ST segment may indicate a condition called myocardial ischemia. However, we must also look at the slope to make any firm decisions. If the slope is upsloping, the likelihood of that person having the condition is lower. If the slope is downsloping, the likelihood of that person having the condition is higher. Neither one of these features by themselves are a predictor of anything, but they may have some predicting power after they are crossed.

b)

We can convert the latitude and longitude into binned values. Each combination of the latitude and longitude bins represents a geographical area of space. The size of the area depends on the size of the bins. In other words, we can split the entire map into a grid, where each square in the grid is represented by latitude and longitude bins. Since location is a good predictor of home price, we can predict the price based on the location of the area of space represented by the two particular bins.

c)

Say we are trying to predict whether or not someone has the common cold. Let us use features that describe the person's symptoms, such as coughing and sneezing. Each of those features alone does not indicate that they have the common cold. For example, one could exhibit sneezing due to allergies or simply feeling cold (not the illness). One could exhibit coughing due to being a smoker. However, if we crossed these two features, there is a high probability that the person is in fact sick with the common cold.

Sample	Sneezing	Coughing	Common Cold
1	Yes	Yes	Yes
2	Yes	No	No
3	No	No	No
4	No	Yes	No
5	Yes	Yes	Yes
6	Yes	No	No
7	No	Yes	No