Study Partner: Kevin Jungho Lee

2 Perceptron Training

Assume a three input perceptron plus bias (it outputs 1 if $b + \sum_i w_i * x_i > 0$, else 0). Assume a learning rate c of 1 and initial weights all 1: $\Delta w_i = c(t - z) * x_i$, where t is the true label and z is the predicted label.

Show weights after each pattern in Table 1 until the result converges. Use an Excel sheet (attach your Excel sheet to the homework). Iterate over the training samples from top to bottom.

x_1	x_2	x_3	t
1	0	1	0
1	1	0	0
1	0	1	1
0	1	1	1

Table 1: Train Set

	w0	w1	w2	w3
Iteration 0	1	1	1	1
Iteration 1	0	0	1	0
Iteration 2	-1	-1	0	0
Iteration 3	0	0	0	1
Iteration 4	0	0	0	1
Iteration 5	-1	-1	0	0
Iteration 6	-1	-1	0	0

Note iteration 2 and iteration 6 have same weights, and the algorithm hasn't converged. Thus, this won't converge, which makes sense because (1, 0, 1) has two different targets, and thus, the dataset is not linearly separable.

3 Input Validation

A SickBit health sensor produces a stream of readings from 20 different sensors (think blood pressure, heart rate body temperature, etc.). List two techniques you could use to check whether the stream of data coming from the sensors are valid or not. Write one or two sentences to describe each approach.

1. Cross Validation: If a sensor is giving values that are very different from 15 other sensors, then the value we are getting from this sensor is probably not valid. The idea is trust the majority.
2. Controlled Testing: Check if the values you get are expected under a controlled environment. For example, if you sit and relax, you would expect to get a reading of normal heart rate from the sensor. If you are running, you would expect to get a high heart rate from the sensor.

4 Distributions

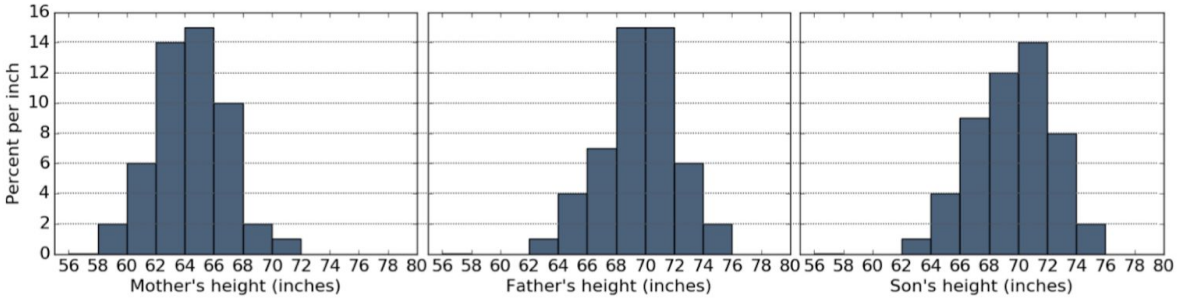


Figure 1: Height Distributions

Galton measured the heights of individuals in 200 families, each of which included one mother, one father, and a varying number of adult sons. The three histograms of heights in Figure 1 depict the distributions for all mothers, fathers, and adult sons. All bars are 2 inches wide. All bar heights are integers. The heights of all people in the data set are included in the histograms.

Since it is difficult to tell the counts immediately by looking at the histogram, let's make some tables that represent the data.

Mother

58-60	60-62	62-64	64-66	66-68	68-70	70-72
4%	12%	28%	30%	20%	4%	2%

Father

62-64	64-66	66-68	68-70	70-72	72-74	74-76
2%	8%	14%	30%	30%	12%	4%

Son

62-64	64-66	66-68	68-70	70-72	72-74	74-76
2%	8%	18%	24%	28%	16%	4%

(a) Calculate each quantity described below or write Unknown if there is not enough information above to express the quantity as a single number (not a range). Show your work!

(i) The percentage of mothers that are at least 60 inches but less than 64 inches tall.

(ii) The percentage of fathers that are at least 64 but less than 67 inches tall.

(iii) The number of sons that are at least 70 inches tall.

(iv) The number of mothers that are at least 60 inches tall.

I. From the table above, we see the percentage of mothers between 60 inches and 64 inches is 40%.

II. We know the percentage of fathers between 64 inches and 66 inches. We also know the percentage of fathers between 66 inches and 68 inches, but not between 66 inches and 67 inches. So, the answer is unknown.

III. We don't know the total number of sons, so the answer is unknown.

IV. The percentage of mothers who are less than 60 inches is 4%. So that is 200 * 0.04 = 8 mothers. So the answer is 200 - 8 = 192.

(b) If the father's histogram were redrawn, replacing the two bins from 72-to-74 and from 74-to-76 with one bin from 72-to-76, what would be the height of its bar? If it's impossible to tell, write Unknown.

The total percentage of this bar will be 16%. Since the height represents "percent per inch", and it is 4 inches, the height will be 16 / 4 = 4.

(c) The percentage of sons that are taller than all of the mothers is between _____ and _____. Fill in the blanks in the previous sentence with the smallest range that can be determined from the histograms, then explain your answer below.

Assume left inclusive. From the diagram, we observe that all mothers are less than 72 inches and the tallest mothers are at least 70 inches. The percentage of sons who are equals to or taller than 72 inches is 20%. The percentage of sons who are equals to or taller than 70 inches are 48%. Thus, the percentage of sons that are taller than all of the mothers is between 20% and 48%.

5 Voronoi

Draw the Voronoi diagram of 10 points all on a line. Draw separately the Voronoi diagram of 10 points all on a circle. What do these two diagrams have in common?

Common Observation: The area of each section is proportional to the distance between the two points.

6 Augmentation

Many methods for making predictions from data, such as linear regression, are limited in terms of the transformations that they can apply to input data before making a prediction. As linear regression assumes that the output is the sum of coefficients multiplied by input features, it is unable to account for cases where the impact of two features together is greater than the sum of their parts. For example, a house that both has > 5 bedrooms and is in California may be worth four times more than would be expected from the learned price impact of each feature on its own.

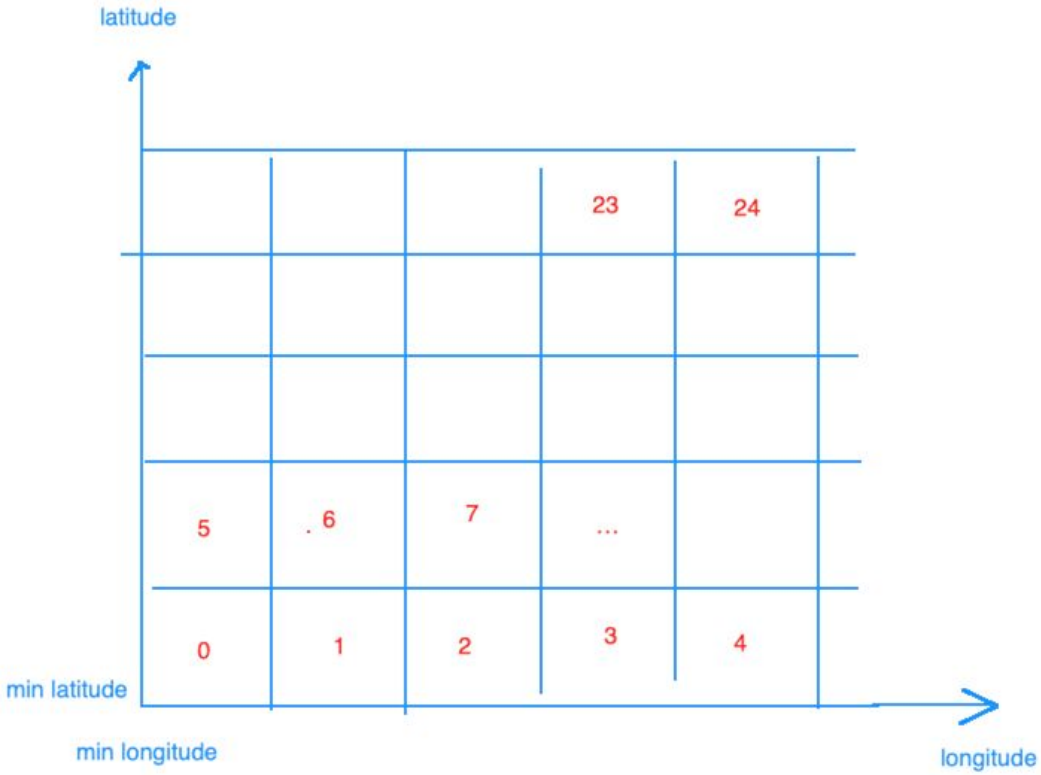
Feature Crosses are synthetic features you can form by crossing two or more features together, and they can help to improve the predictive power of techniques such as linear regression. Expanding on the above housing example, you could generate a new feature that indicates a combination of both a home's number of bedrooms and location.

(a) Describe two pairs of features from Project 2 that might be interesting to cross together, and explain why.

I think it would be interesting to cross sex and fbs (Fasting blood sugar > 120 mg/dl (1 = true; 0 = false)). So we can do something like creating a new feature, where 1 means man and fbs is 1, 0 means otherwise. This is because maybe being male and having high blood sugar is very dangerous, but it's not if you are female. This is my random theory, but maybe useful, you never know.

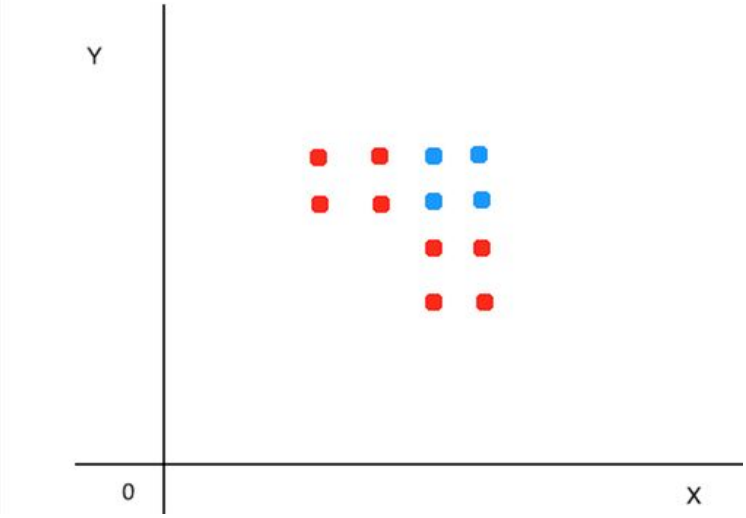
(b) You have latitude and longitude for homes, and you think feature crosses may allow you to make better predictions. However, your latitude and longitude are continuously valued. How might you do a feature cross in this case?

First of all, it's useful to feature cross latitude and longitude because the combination of these two features gives us "neighborhood" feature. Now, to create this new feature, you can bucket latitude and longitude into, say 5 buckets respectively. Then you can assign numbers to each block (numbered 0 - 24). So, when you feature cross latitude and longitude, you get a new feature numbered 0 - 24, where each number corresponds to an area. (Note that we want to one hot encode this feature because it is categorical where the order doesn't matter).



(c) Think up a dataset consisting of features X and Y and associated labels Z that is shaped such that a linear model would perform poorly without feature crosses. Provide a table with at least 7 points from your dataset.

Consider a dataset with two features X and Y as follows:



X	Y	Target Label
4	6	0 (red)
4	7	0 (red)
5	6	0 (red)
5	7	0 (red)
6	4	0 (red)
6	5	0 (red)
7	4	0 (red)
7	5	0 (red)
6	6	1 (blue)
6	7	1 (blue)
7	6	1 (blue)
7	7	1 (blue)

Clearly from the drawing, this is not linearly separable. But, we can do feature crosses! We can feature cross X and Y and create a new feature Z, where Z is 1 if X >= 6 and Y >= 6. Z is 0 otherwise. Then we can obtain a new table as follows:

X	Y	Z	Target Label
4	6	0	0 (red)
4	7	0	0 (red)
5	6	0	0 (red)
5	7	0	0 (red)
6	4	0	0 (red)
6	5	0	0 (red)
7	4	0	0 (red)
7	5	0	0 (red)
6	6	0	1 (blue)
6	7	0	1 (blue)
7	6	1	1 (blue)
7	7	1	1 (blue)

Now our learning model can just say, “Target is 1 if Z is 1. Target is 0 otherwise”.