CS 188 HW 2 Solutions

Due on Friday, February 14 at 11:59PM

1 Instructions:

You may form small groups (e.g. of up to four people) to work on this assignment, but you must write up all solutions by yourself. List your study partners for the homework on the first page, or "none" if you had no partners.

Keep all responses brief, a few sentences at most. Show all work for full credit.

Start each problem on a new page, and be sure to clearly label where each problem and subproblem begins. All problems must be submitted in order (all of P1 before P2, etc.).

No late homeworks will be accepted. This is not out of a desire to be harsh, but rather out of fairness to all students in this large course.

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

Solution: Showing all of the math would be a bit heavy, so instead we show how the weights changed with each iteration as well as the results of a single iteration. The input data is not linearly separable, and as such the weights will never converge. As such, we need to recognize this and stop iteration early. Bias can be treated as a weight w_0 for a new feature x_0 that is always valued at one.

For the first sample, we can calculate our forward pass as

$$b + \sum_i w_i * x_i$$

which is equivalent to

$$b + w_1 x_1 + w_2 x_2 + w_3 x_3 = 1 + 1 * 1 + 1 * 0 + 1 * 1 = 3$$

As our output value is > 0, our perceptron outputs z as 1. We now update weights as follows

$$\Delta w_i = c(t - z) * x_i = \Delta w_i = 1(0 - 1) * x_i = -x_i$$

for each weight from 0 to 3 (weight 0 is b). What this weight update rule means is that for each feature in our input sample that was 1, we will subtract 1 from the corresponding weight (including bias) and for each feature that was 0, no change will be made. For example:

$$w_1 = w_{1\_old} + \Delta w_1 = w_{1\_old} + c(t - z) * x_1 = w_{1\_old} + 1(0 - 1) * x_1 = 1 - x_1 = 1 - 1 = 0$$

Thus, we decrement b, w_1, and w_3 by one, giving us the new weights seen in the row of Table 2 corresponding to iteration one.

Iteration	x_1	x_2	x_3	t	z	b	w_1	w_2	w_3
0	n/a	n/a	n/a	n/a	n/a	1	1	1	1
1	1	0	1	0	1	0	0	1	0
2	1	1	0	0	1	-1	-1	0	0
3	1	0	1	1	0	0	0	0	1
4	0	1	1	1	1	0	0	0	1
5	1	0	1	0	1	-1	-1	0	0
6	1	1	0	0	0	-1	-1	0	0
7	1	0	1	1	0	0	0	0	1

Table 2: Parameters after each update

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 streams of data coming from the sensors are valid or not. Write one or two sentences to describe each approach.

Solution: There are a wide range of valid solutions to this question. The following are good examples:

- Ensure that all sensor readings are within a predefined range, e.g. $25 <$ `Heart Rate` < 220.
- Ensure that sensor readings are present, e.g. `Temp` $\neq NaN$
- Ensure that the values you're seeing in the real world are similar (mean, standard deviation, etc.) to the values you trained on.
- Ensure that there is no drift in the sensors, e.g. today's mean body temperature is similar to that from two weeks ago

4 Distributions

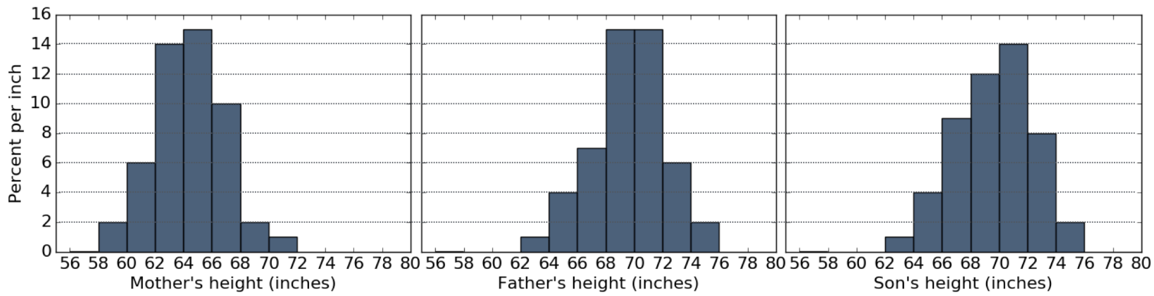

Figure 1: Height Distributions

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.

(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.

Solution: (60-64): 2 inches * (6 + 14) percent/inch = **40 percent**

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

Solution: *Unknown*: We can't tell how heights are distributed within a bin.

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

Solution: *Unknown*: The total number of sons is unknown, so the size of any subset is unknown.

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

Solution: (100 percent - (2 inches * 2 percent/inch)) * 200 mothers = **192** mothers

(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.

Solution: The bin contains 6 * 2 + 2 * 2 = 16 percent, and the width is 4 inches, so the height is **4 percent/inch.**

(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.

Solution: The tallest mother is between 70 and 72 inches. The proportion of sons above 72 inches is (8 + 2) percent/inch * 2 inches = 20 percent. The proportion of sons above 70 inches is (14 + 8 + 2) percent/inch * 2 inches = 48 percent.

So, 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?

Solution: Most loose similarities would be accepted here. Some common ones may include both diagrams having 10 segments or both diagrams having segments of infinite area.

One diagram should look like a sliced up pizza (triangular wedges) and the other should look like adjacent rectangles. Each should have 10 segments.

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.

Solution: Any two features that sound like they may have clinical relevance together should be good. For example, age and sex or cholserum and fasting blood sugar.

(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?

Solution: Directly crossing latitude and longitude while they remain continuously valued would result in a new bucket for each unique lat,lon pair you see, resulting in a potentially infinite number of categories. As such, we need a way to restrict the number of categories generated. One such approach is to bin latitude and longitude, changing them from continuously valued to categorical. The scale of these bins may vary depending on the problem at hand, such as city scale or nation scale. As an example, 22.135 could simply be binned to 20-30. This greatly reduces the number of values possible in the input features, correspondingly keeping the number of categories resulting from the cross manageable.

tl;dr: Binning. Don't accept answers about multiplication.

(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.

Solution: There's a good deal of flexibility here, but all responses must have at least 7 datapoints and contain 2 feature columns XY and one label column Z. The datapoints should be such that there is no linear decision boundary, but a feature cross allows for significantly better performance. One acceptable solution is below in table 3. This dataset exhibits a nonlinear effect when both X and Y are true.

X	Y	Z
1	0	1
1	0	1
1	0	1
0	1	.5
0	1	.5
1	1	20
1	1	20

Table 3: Dataset needing feature crosses