1. (a) $P_c = 0.7$ $P_a = 0.2$ $P_s = 0.1$

$H = -\sum_k P_k \log P_k = -0.7\log 0.7 - 0.2\log 0.2 - 0.1\log 0.1 = 1.16$

(b) $H_{max} = 3\times(-\frac{1}{3}\log\frac{1}{3}) = 1.58$

(c) $H_{parent} = 1.16$

For Gender $H_{child1} = -0.8\log 0.8 - 0.12\log 0.12 - 0.08\log 0.08 = 0.916$

$H_{child2} = 1.218$

$Gain = 1.16 - \frac{1}{4}\times 0.916 - \frac{3}{4}\times 1.218 = 0.0175$

For StudentType $Gain = 1.16 - \frac{1}{2}H(0.76, 0.16, 0.08) - \frac{1}{2}H(0.64, 0.24, 0.12)$

$= 1.16 - \frac{1}{2}\times 1.015 - \frac{1}{2}\times 1.273$

$= 0.016$

Split on Gender would get better information gain for first decision

2. (a) $w\cdot x = -\frac{3}{7} + \sum_{i}^{7} \frac{x_i}{7}$. $w = [-\frac{3}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}]$

(b) threshold function $h_w(x) = \sin wx$ $w = [\frac{\pi}{2}, \frac{\pi}{2}, \frac{\pi}{2}, \frac{\pi}{2}]$

(c) Input layer $w_1 = [10, -1, -1]$ $wx = 10 - x_1 - x_2$ $g_1(wx)$: if $wx \geq 0$, output -1, if $wx < 0$, output 1

~~Hidden layer~~ $w_2 = [0, 1, 0]$, $w_3 = [0, 0, 1]$ $w_2x = x_1$, $w_3x = x_2$

$g_2(w_2x)$: if $w_2x > 0$: output 1; if $w_2x < 0$; output -1 $w_4 = [-1, 1, 1]$ $g_2(-g_1(w_1x) + g_2(w_2x) + g_2(w_3x))$

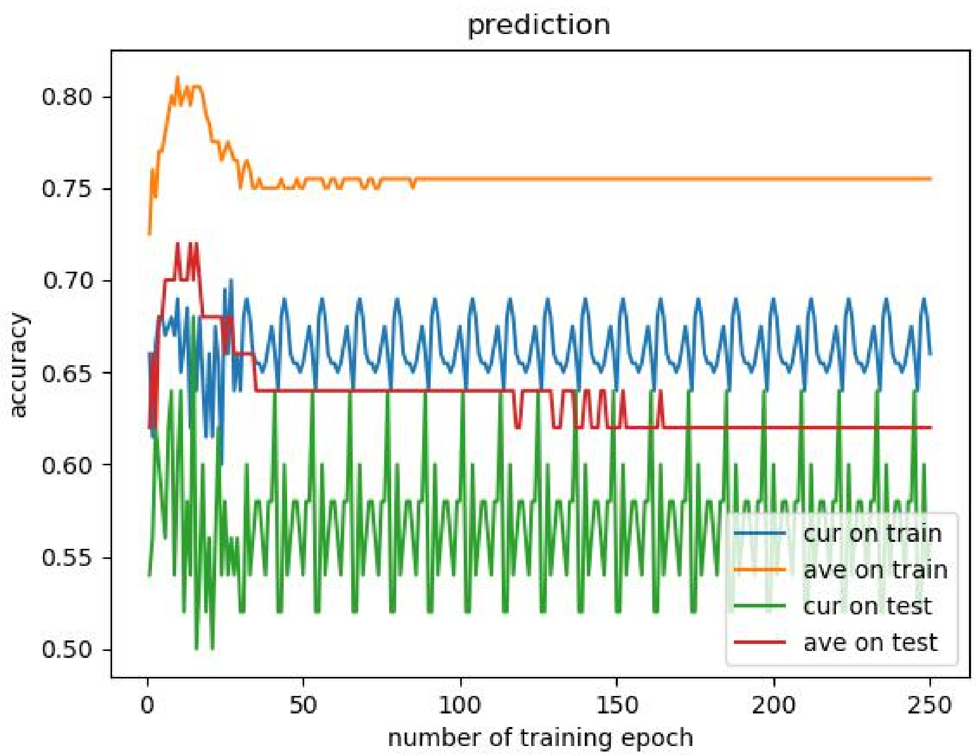
3. majority classifier predict score about 50% without leave one out. Leave one out result in unbalance of how the class distributes in the data set, if leave one positive, the majority is negative. So we can't use leave one out here.

4.

Instruction command: python hw3a.py

(a) If we encode the categorical data into numerical numbers, like A,B,C,D to 0,1,2,3. This may mislead the model because the number implies some kind of order or hierarchy which actually is not. In the perceptron algorithm we finally get a average weight and then dot multiply it with vector x to predict. So we can not set (mood: silly, happy, tired) to (0,1,2) because the value will affect how the algorithm learn weight and mislead.

(b)



Average model works better than current model both on train set and test set

c)

after 10 training epoch the best model is

[0.5155 1.1895 -1.334 0.66 0.081 0.263 0.1715 0.166 -0.077 0.4265 -2.6725 3.5735 -0.3855 -0.371 1.0865 0.2455 -0.7065]

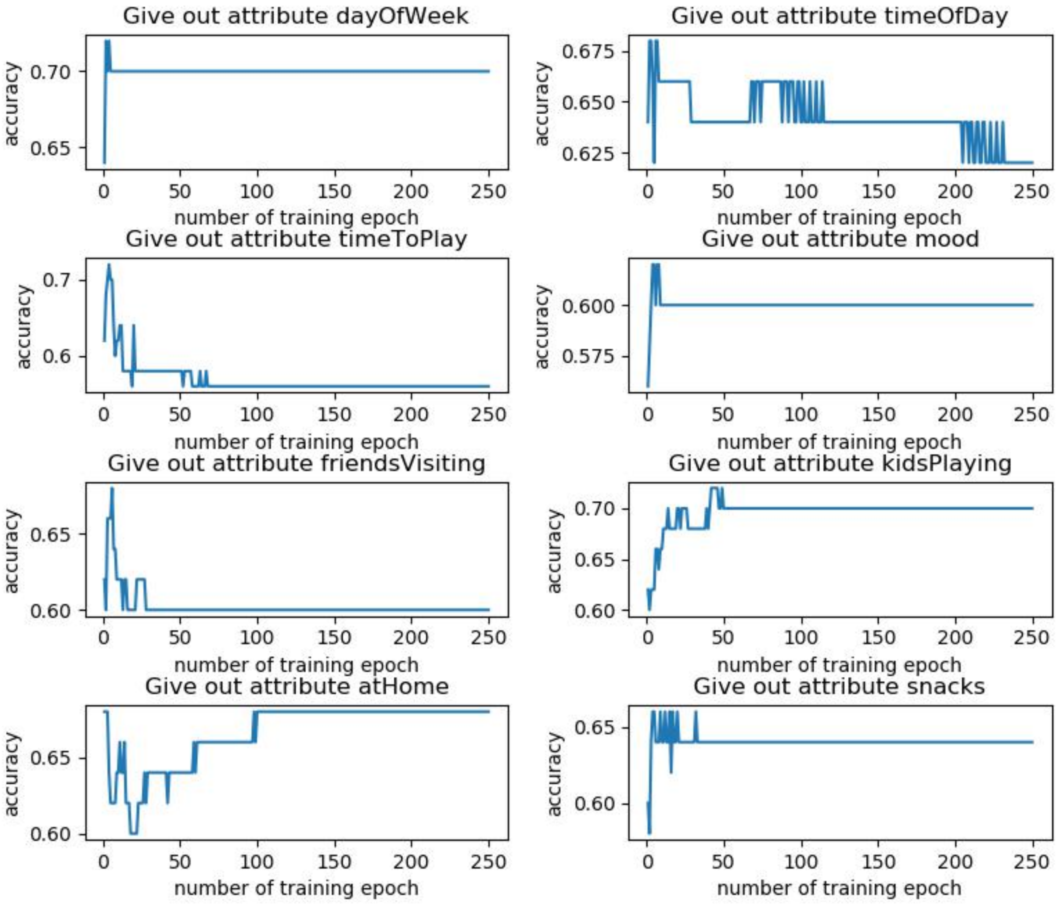
so the math description of decision function is

w • X = 0.5155 +1.1895 -1.334dayOfWeek +0.66dayOfWeek +0.081dayOfWeek +0.263timeOfDay +0.1715timeOfDay +0.166timeOfDay -0.077timeToPlay +0.4265timeToPlay -2.6725timeToPlay +3.5735mood -0.3855friendsVisiting -0.371kidsPlaying +1.0865atHome +0.2455snacks -0.7065game

where threshold is 0, if bigger than threshold then predict value is SettersOfCatan, and otherwise

X = [1,'Weekday', 'Saturday', 'Sunday', 'morning', 'afternoon', 'evening','<30', '30-60', '>60', 'silly', 'happy', 'tired', 'friendsVisiting','kidsPlaying', 'atHome', 'snacks'] the attribute mood plays the most important role that have coeffecient 3.5735

(d)



without attribute ['dayOfWeek'] accuracy: 0.7

without attribute ['timeOfDay'] accuracy: 0.62

without attribute ['timeToPlay'] accuracy: 0.56

without attribute ['mood'] accuracy: 0.6

without attribute ['friendsVisiting'] accuracy: 0.6

without attribute ['kidsPlaying'] accuracy: 0.68

without attribute ['atHome'] accuracy: 0.68

without attribute ['snacks'] accuracy: 0.64

The attribute timeToPlay seems to play the most important role

(e) Examine the weights is better because it's more easy to find out the most important attribute than ablation test, with more efficiency and accuracy. Also if the weights are close on the influential attributes it's hard to figure out just by ablation test. Simply comparing different accuracy by ablation test is not fair because with different attributes the models are different.

(f) The averaged model should be better because it generalize better to test data than the final train model.