Name: Karan Sikka (@ksikka)

15-381 F13

Due: 11/5/13

1) $H(V) = - \left(\frac{\#pos}{total}\right) \cdot \log \left(\frac{\#pos}{total}\right) - \ldots$

So, 15 vaccines, 11 not, 26 total =>

$H(V) = - \left(\frac{15}{26}\right) \lg \left(\frac{15}{26}\right) - \left(\frac{11}{26}\right) \lg_2 \left(\frac{11}{26}\right) = \underline{0.983}$

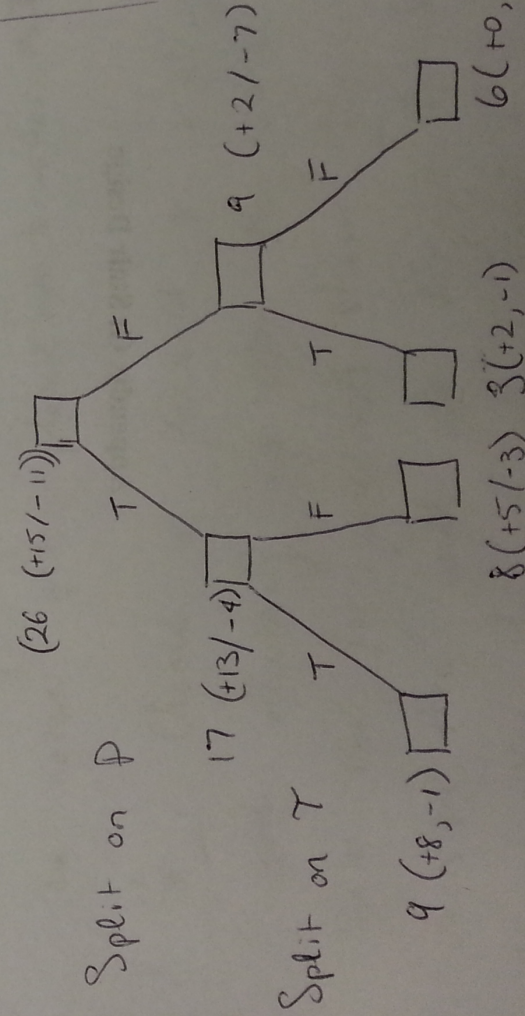2) $IG(P) = H(V) - H(V|P)$

$H(V)$ computed above, what is $H(V|P)$?

$H(V|P) = \frac{17}{26} \left[-\frac{13}{7} \lg_2 \left(\frac{13}{7}\right) - \frac{4}{17} \lg_2 \left(\frac{4}{17}\right) \right] + \frac{9}{26} \left[-\frac{2}{9} \lg_2 \left(\frac{2}{9}\right) - \frac{7}{9} \lg_2 \left(\frac{7}{9}\right) \right]$

$\Rightarrow H(V|P) = 0.779 \quad \Rightarrow IG(P) = \underline{0.204}$

By similar computation,

$H(V|T) = 0.806 \quad \Rightarrow IG(T) = \underline{0.177}$

3) We split in order of largest info gain.

Split on P

$(26 (+15/-11))$

$17 (+13/-4)$ $9 (+2/-7)$

T   F

Split on T

$9 (+8,-1)$ $8(+5(-3)$ $3(+2,-1)$ $6(+0,-6)$

T   F   T   F

Key:
$A (+B, -C)$
=> $A = \#$ of data points
$B = \#$ of positive points
$C = \#$ of negative points

4) Collecting unimportant features would likely lead to overfitting, since the algorithm would think this data is more significant than it actually is.

Making paleness a continuous attribute would also provide the algorithm with unnecessary information to branch on, leading to overfitting.

5) Stopping growth of the tree is a good way to reduce overfitting.

(Less likely to fit noise). This is because noise will have less information gain, and they will not get fit, rather the noise gets placed into broader data.

6) Yes. Choose a threshold x and y which effectively partition the data into the for quadrants shown. The decision tree could use discrete thresholds to partition the continuous space. A quadrants require two levels.