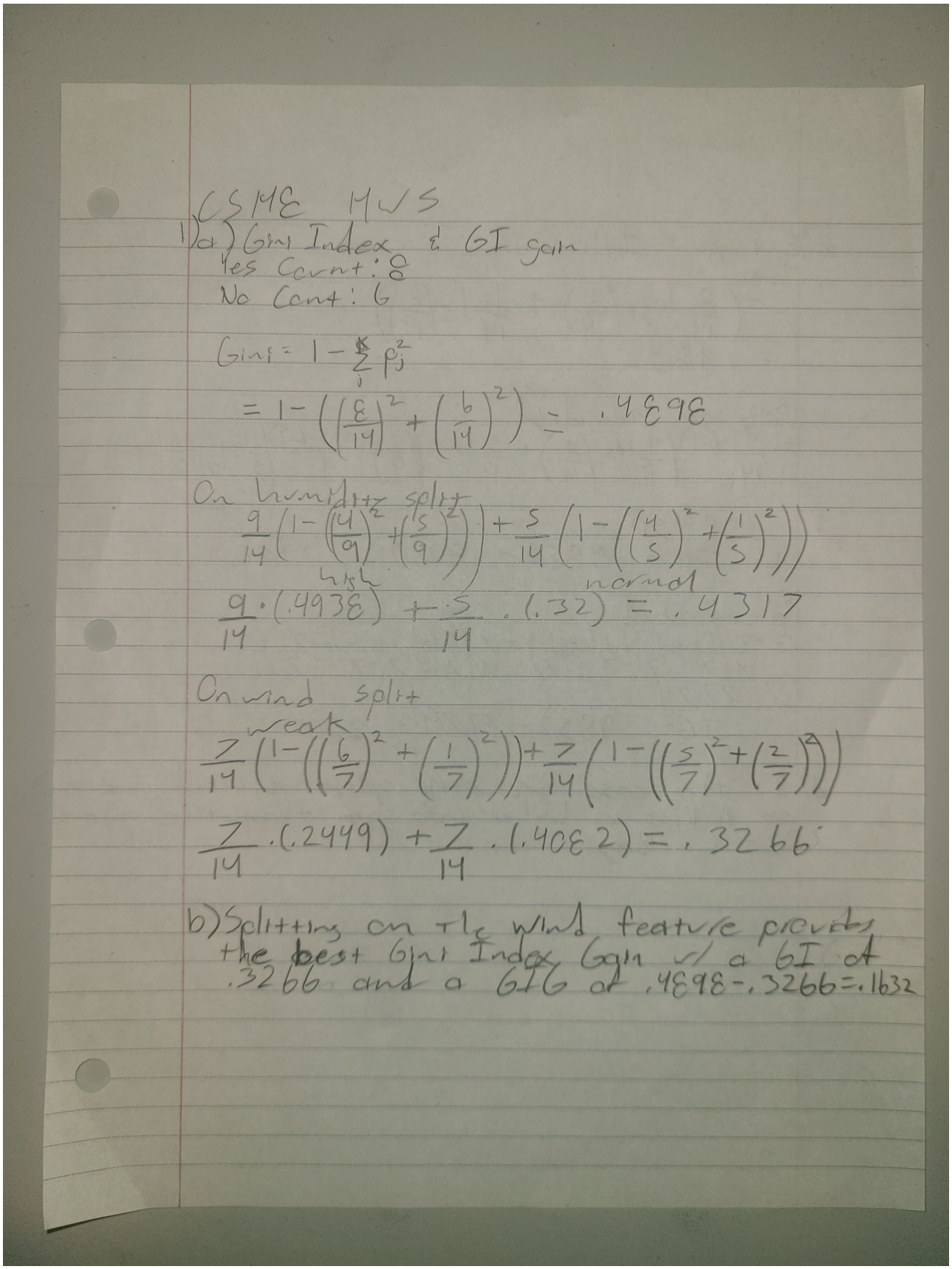


CSME HW5

1) a) Gini Index & GI gain

Yes Count: 8

No Cont: 6

$Gini = 1 - \sum_j^{\#} p_j^2$

$= 1 - \left(\left(\frac{8}{14}\right)^2 + \left(\frac{6}{14}\right)^2\right) = .4898$

On humidity split

$\frac{9}{14}\left(1 - \left(\left(\frac{4}{9}\right)^2 + \left(\frac{5}{9}\right)^2\right)\right) + \frac{5}{14}\left(1 - \left(\left(\frac{4}{5}\right)^2 + \left(\frac{1}{5}\right)^2\right)\right)$

high normal

$\frac{9}{14} \cdot (.4938) + \frac{5}{14} \cdot (.32) = .4317$

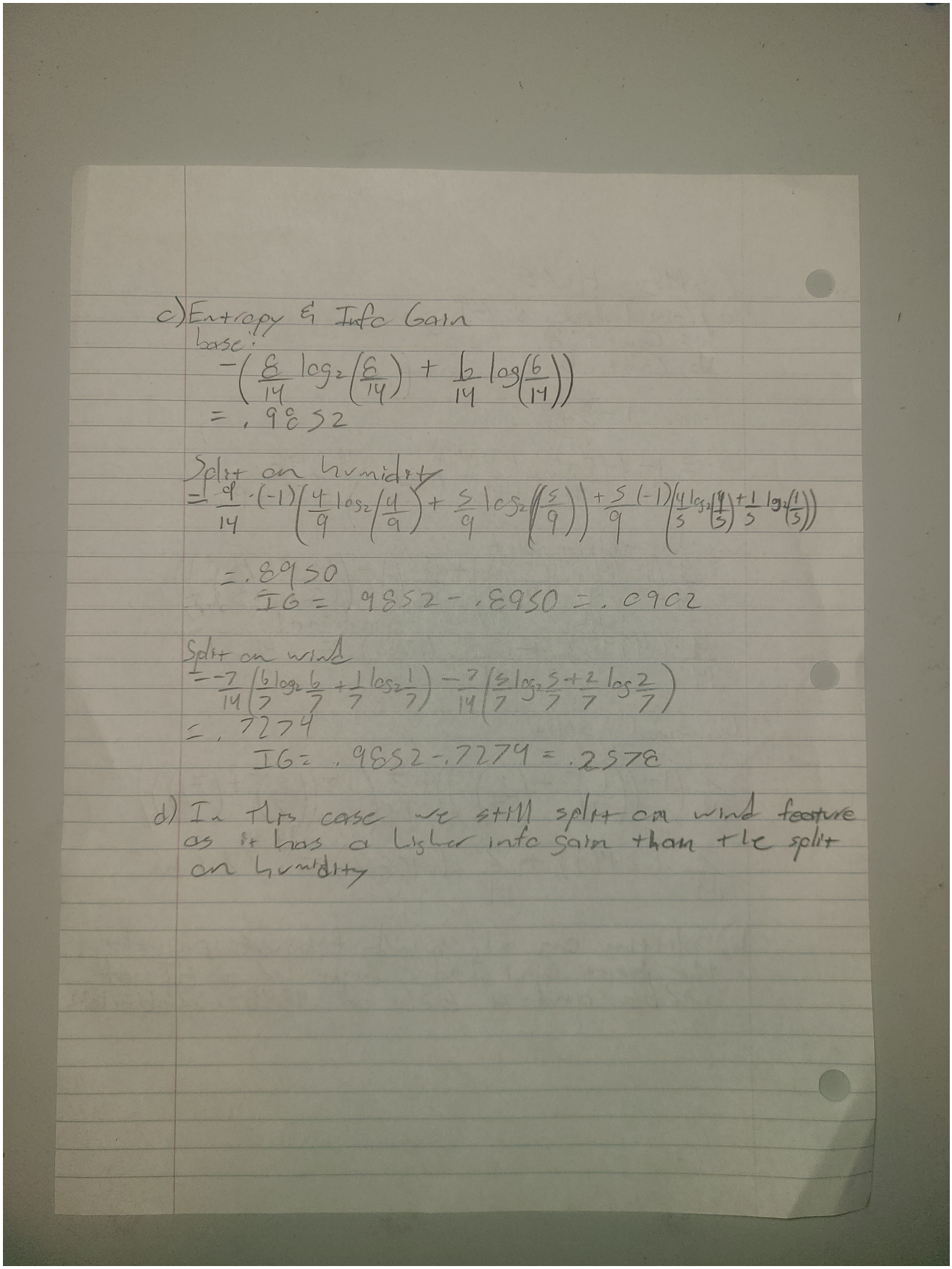
On wind split

weak

$\frac{7}{14}\left(1 - \left(\left(\frac{6}{7}\right)^2 + \left(\frac{1}{7}\right)^2\right)\right) + \frac{7}{14}\left(1 - \left(\left(\frac{5}{7}\right)^2 + \left(\frac{2}{7}\right)^2\right)\right)$

$\frac{7}{14} \cdot (.2449) + \frac{7}{14} \cdot (.4082) = .3266$

b) Splitting on the wind feature provides the best Gini Index Gain w/ a GI of .3266 and a GIG of .4898 - .3266 = .1632



c) Entropy & Info Gain

base:

$$-\left(\frac{6}{14}\log_2\left(\frac{6}{14}\right)+\frac{8}{14}\log\left(\frac{8}{14}\right)\right)$$

$= .9852$

Split on humidity

$$=\frac{9}{14}\cdot(-1)\left(\frac{4}{9}\log_2\left(\frac{4}{9}\right)+\frac{5}{9}\log_2\left(\left(\frac{5}{9}\right)\right)+\frac{5}{9}(-1)\left(\frac{4}{5}\log_2\left(\frac{4}{5}\right)+\frac{1}{5}\log_2\left(\frac{1}{5}\right)\right)\right.$$

$= .8950$

$IG = .9852 - .8950 = .0902$

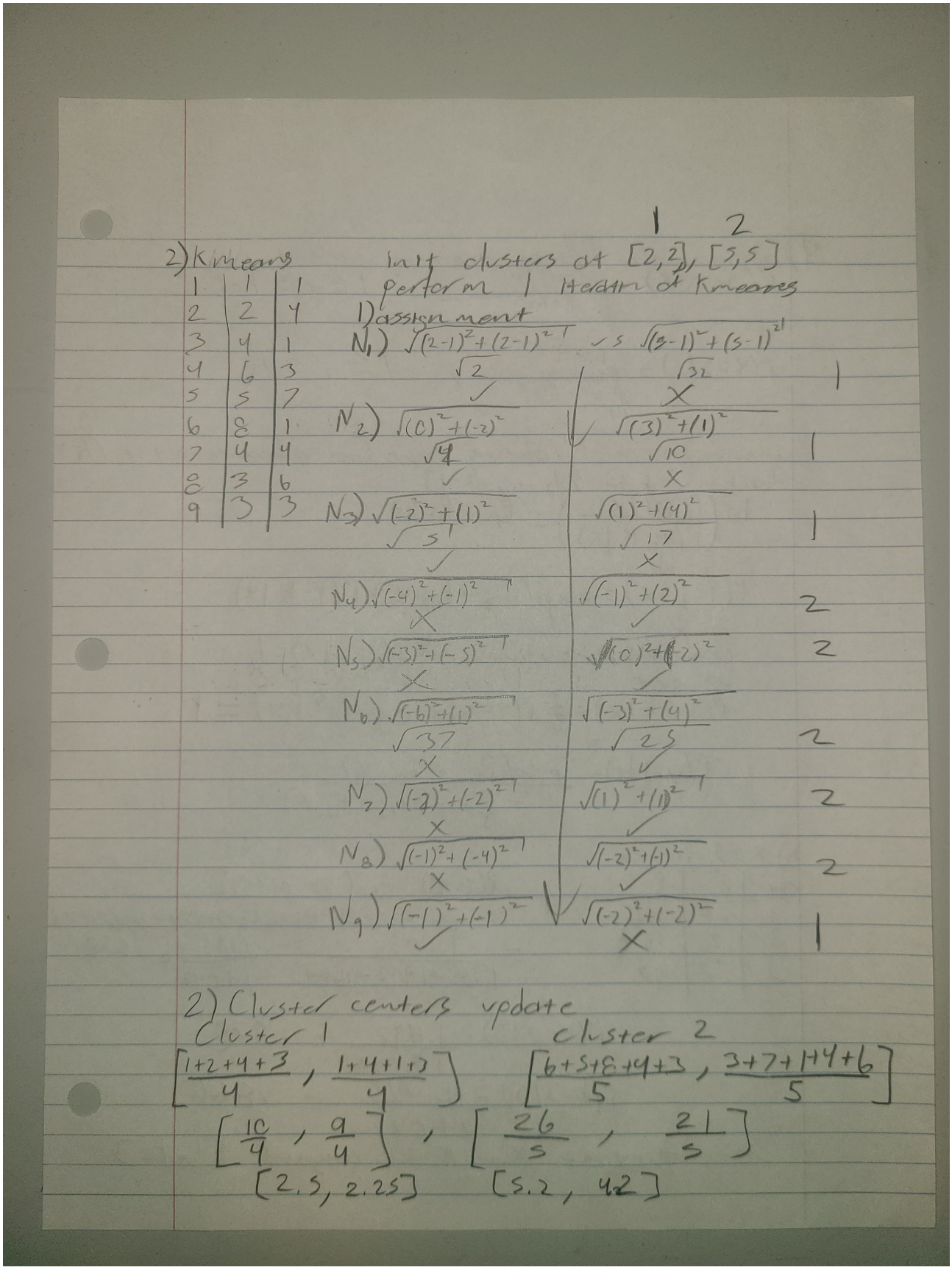
Split on wind

$$=\frac{-7}{14}\left(\frac{6}{7}\log_2\frac{6}{7}+\frac{1}{7}\log_2\frac{1}{7}\right)-\frac{7}{14}\left(\frac{5}{7}\log_2\frac{5}{7}+\frac{2}{7}\log\frac{2}{7}\right)$$

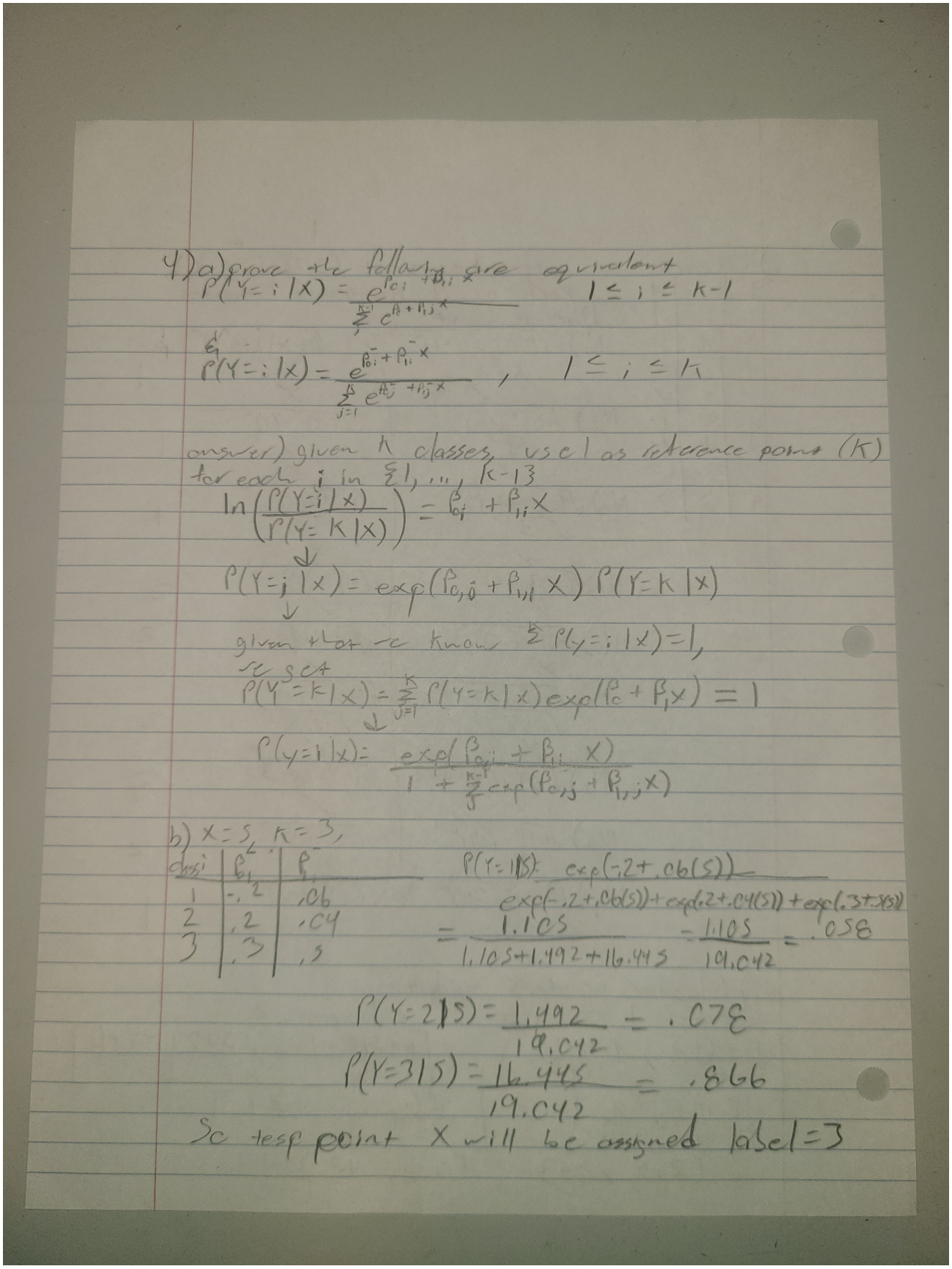
$= .7274$

$IG = .9852 - .7274 = .2578$

d) In this case we still split on wind feature as it has a higher info gain than the split on humidity



3) a)Given that none of the non-circled points are support vectors, removing them will not affect the decision bounday of the SVM b)A hard margin SVM does not allow for any misclassified points. Soft margin SVMs do allow for some incorrect points and also take into account points within the margin to still maximize the margin. c)If we remove the + sample that is circled the number of support vectors becomes 5. The new line seperating the two would be a line equidistant from the left 3 +'s and the two circled -'s.



4) a) prove the following are equivalent

$$P(Y=i|X) = \frac{e^{\beta_{0i}+\beta_{1i}X}}{\sum_{j=1}^{K-1} e^{\beta_{0j}+\beta_{1j}X}} \qquad 1 \le i \le K-1$$

$$P(Y=i|X) = \frac{e^{\beta_{0i}+\beta_{1i}X}}{\sum_{j=1}^{K} e^{\beta_{0j}+\beta_{1j}X}}, \qquad 1 \le i \le K$$

answer) given K classes, use class as reference point (K)

for each i in $\{1, \ldots, K-1\}$

$$\ln\left(\frac{P(Y=i|X)}{P(Y=K|X)}\right) = \beta_{0i} + \beta_{1i}X$$

$$P(Y=i|X) = \exp(\beta_{0,i} + \beta_{1,i}X)\, P(Y=K|X)$$

given that we know $\sum P(y=i|X)=1$,

we get

$$P(Y=K|X) = \sum_{j=1}^{K} P(Y=K|X)\exp(\beta_0 + \beta_1 X) = 1$$

$$P(y=i|X) = \frac{\exp(\beta_{0i} + \beta_{1i}X)}{1 + \sum_{j=1}^{K-1}\exp(\beta_{0,j} + \beta_{1,j}X)}$$

b) $X=5$, $K=3$,

class i	β_0	β_1
1	-.2	.06
2	.2	.04
3	.3	.5

$$P(Y=1|5) = \frac{\exp(-.2+.06(5))}{\exp(-.2+.06(5))+\exp(.2+.04(5))+\exp(.3+.5(5))} = \frac{1.105}{1.105+1.492+16.445} = \frac{1.105}{19.042} = .058$$

$$P(Y=2|5) = \frac{1.492}{19.042} = .078$$

$$P(Y=3|5) = \frac{16.445}{19.042} = .866$$

So test point X will be assigned label=3

5) a)False, for regression trees we pick the feature whose split minimizes the MSE b)False, K-means finds the local optima for its clusters so its final cluster values with depend on the initial conditions. c)True, for aggomerative clustering we begin with k=n where k is equal to the number of clusters and n is the number of points in the dataset. We then combine clusters in a way that minimizes the tradeoff between the number of clusters and distortion. d)False, a smaller lambda provides a greater margin as it penalizes error. e)True, we use a random forest to prevent overfitting as it allows us to take the average or majority vote of multiple shallow trees.