Kai Wong (704451679)

CS 145 Fall 2018

Homework 2

**1.1:** **Decision Tree Construction**

Let class P: Democrat, class N: Republican

|  |  |  |  |
| --- | --- | --- | --- |
| Vote for handicapped-infants | pi | ni | I(pi, ni) |
| Y | 6 | 2 | 0.811 |
| N | 4 | 8 | 0.918 |

|  |  |  |  |
| --- | --- | --- | --- |
| Vote for water-project | pi | ni | I(pi, ni) |
| Y | 4 | 6 | 0.971 |
| N | 6 | 4 | 0.971 |

|  |  |  |  |
| --- | --- | --- | --- |
| Vote for budget-resolution | pi | ni | I(pi, ni) |
| Y | 9 | 2 | 0.684 |
| N | 1 | 8 | 0.503 |

Info(D) = = 1.0

Info­handicapped(D) = = 0.875

Info­water-project(D) = = 0.971

Info­budget-resolution(D) = = 0.603

Gain(handicapped) = Info(D) - Info­handicapped(D) = 0.125

Gain(water-project) = Info(D) - Info­water-project(D) = 0.029

Gain(budget-resolution) = Info(D) - Info­budget-resolution(D) = 0.397

Info(Dbudget-resolution=Y) = = 0.684

Infohandicapped(Dbudget-resolution=Y) = 3,1) = 0.672

Infowater-project(Dbudget-resolution=Y) = 6,1) = 0.672

Gainbudget-resolution=Y(handicapped) = 0.012

Gainbudget-resolution=Y(water-project) = 0.012

Info(Dbudget-resolution=N) = = 0.503

Infohandicapped(Dbudget-resolution=N) = 1,7) = 0.483

Infowater-project(Dbudget-resolution=N) = 0,3) = 0.433

Gainbudget-resolution=N(handicapped) = 0.020

Gainbudget-resolution=N(water-project) = 0.070

This is the resulting decision tree using information gain as the attribute selection heuristic as computed above. This tree is pruned (when all leaf nodes of a parent node evaluate to the same value V, subtree starting from parent node is pruned from main tree and parent node is replaced with V).

Tree pruning was performed for the following:

**Budget -> N -> Republican:**

Budget -> N -> Water -> N -> Republican

Budget -> N -> Water -> Y -> Handicap -> Y -> Republican

Budget -> N -> Water -> Y -> Handicap -> N -> Republican (By majority voting)

**Budget -> Y -> Handicap -> Y -> Democrat:**

Budget -> Y -> Handicap -> Y -> Water -> Y -> Democrat

Budget -> Y -> Handicap -> Y -> Water -> N -> Democrat (By majority voting)

Vote for budget-resolution-adoption

Vote for handicapped-infants

Y

Y

N

N

Republican

Democrat

Vote for water-project-cost-sharing

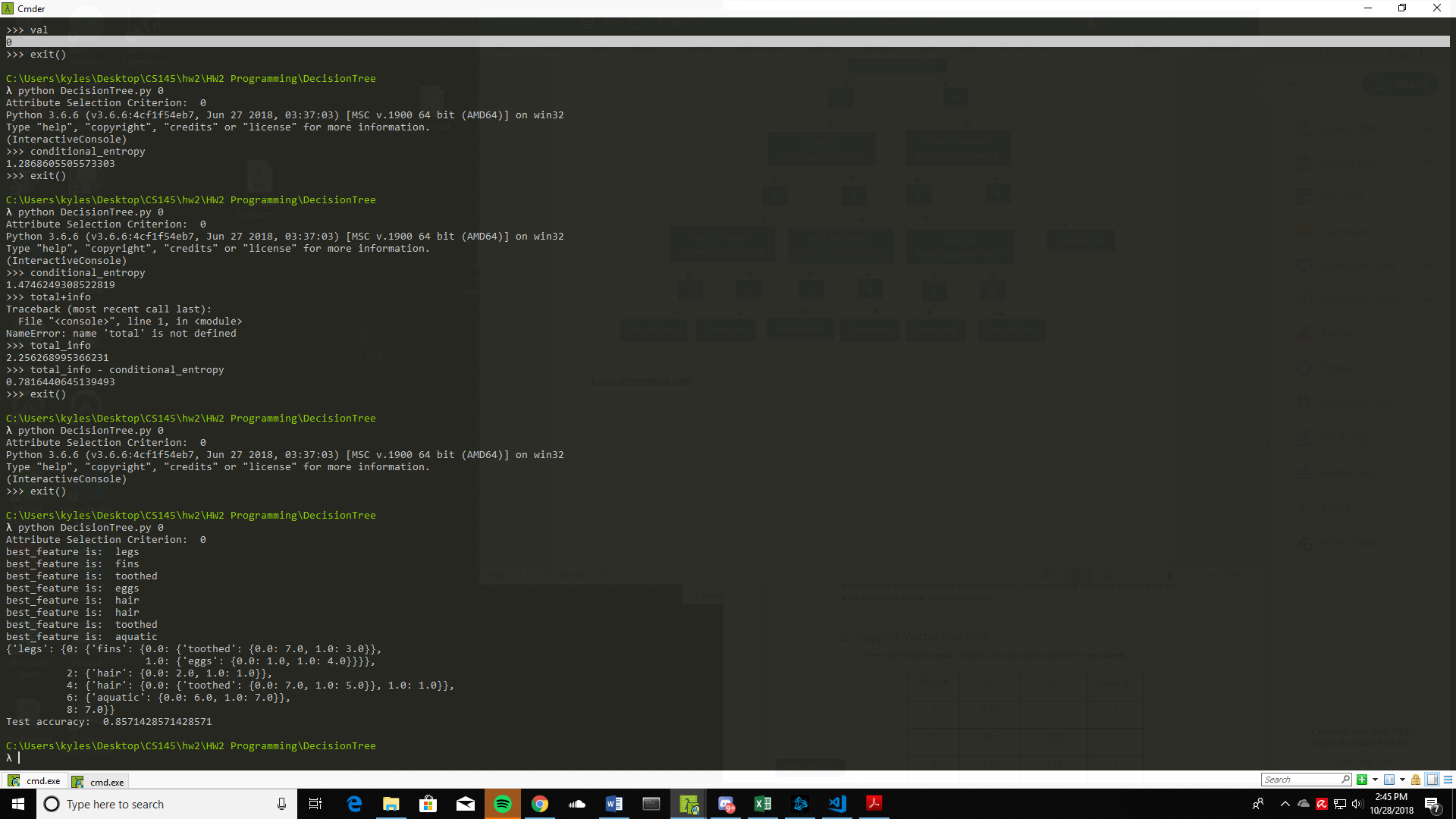
Y

N

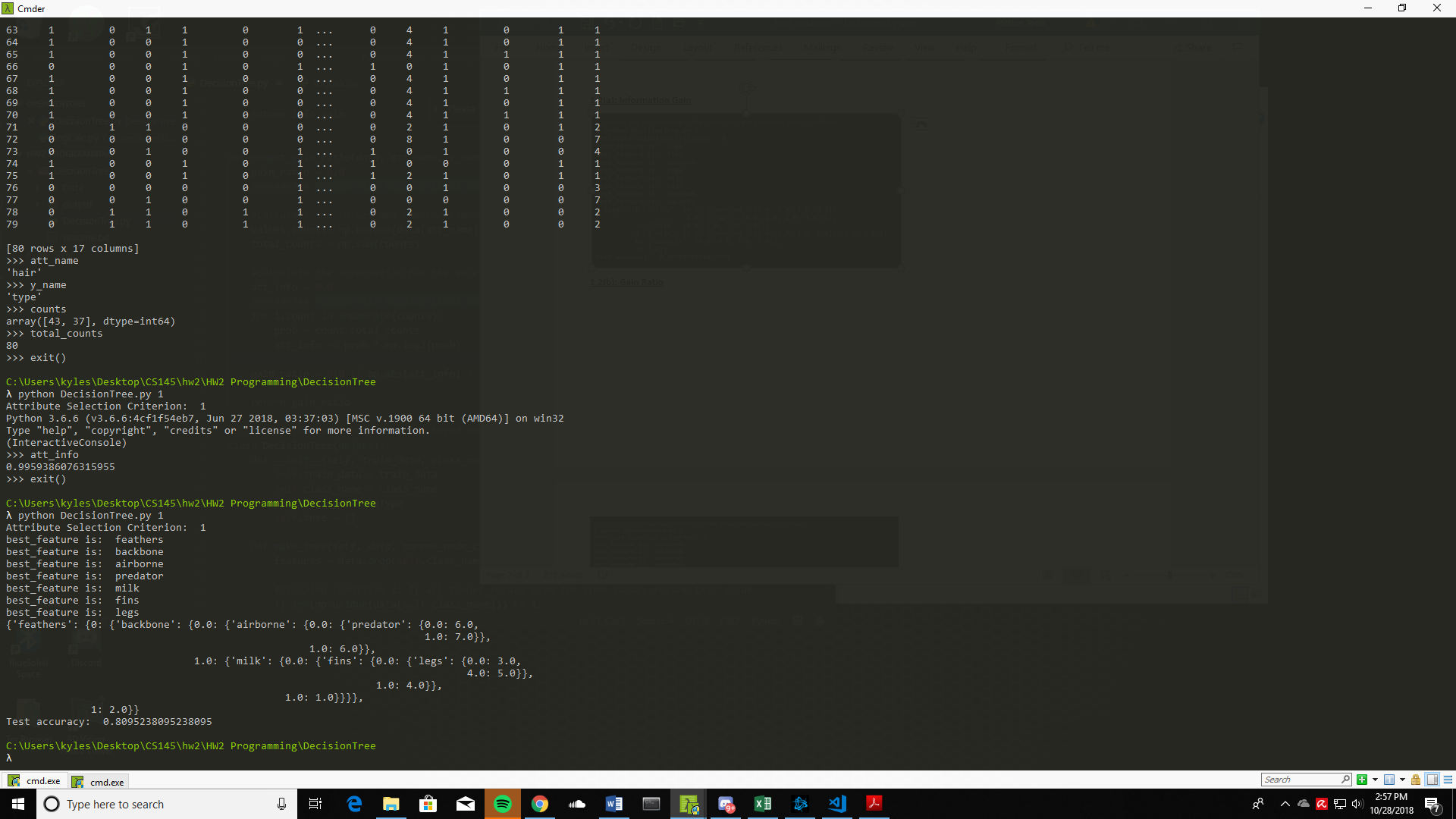
Republican

Democrat

**1.2(a): Information Gain**



**1.2(b): Gain Ratio**



The accuracy when using information gain as the attribute selection heuristic was 0.857 whereas using gain ratio gave an accuracy of 0.810. A possible reason for the decreased performance when using gain ratio is because while information gain is typically biased towards multivalued attributes, almost all of the attributes in the dataset were binary except for ‘legs’, hence it would not be necessary to apply gain ratio to overcome these biases from using information gain.

Furthermore, gain ratio tends to prefer unbalanced splits in which one partition is much smaller than the others (since smaller partitions would have a smaller SplitInfo => larger gain ratio). As a result, it may overcompensate and choose an attribute just because its intrinsic information is low despite having lower than average information gain. This can be seen in the feature selection choice: information gain first selected ‘legs’ as the best feature whereas gain ratio selected ‘feathers’ when intuitively, it would not be hard to imagine why partitioning animals based on their number of legs first would be a better choice over if they had feathers or not. Hence, information gain would be a better attribute selection measure than gain ratio for this dataset.

**2.1(a):** **Please point out the support vectors in the training points**

Each non-zero αi indicates that the corresponding xi is a support vector.

Hence, the support vectors are points 2, 6, and 18

**2.1(b):** **Calculate the normal vector of the hyperplane**

**2.1(c):** **Calculate the bias b**

**2.1(d):** **Write the learned decision boundary function**

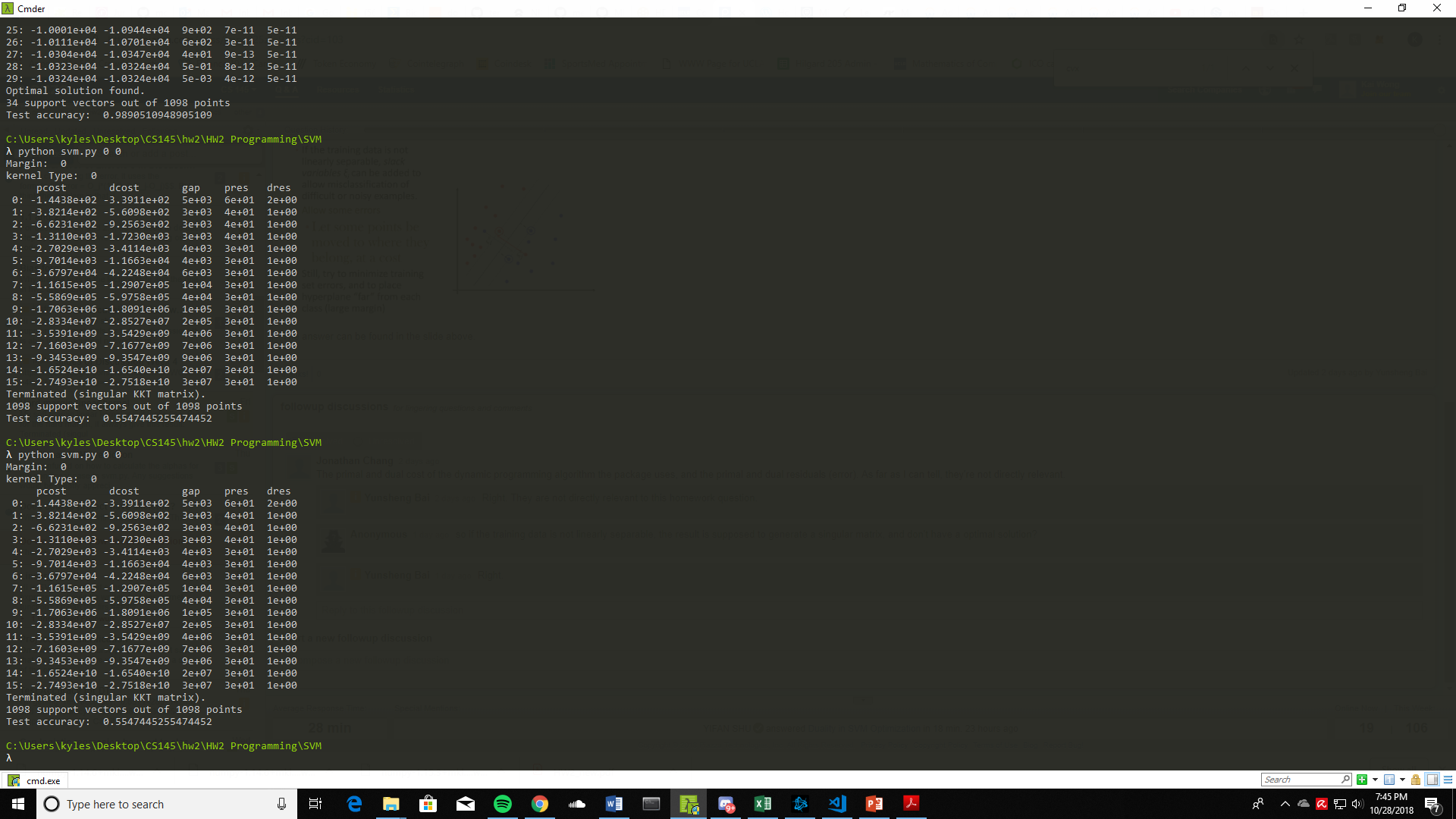
**2.1(e):** **Predict class label of new data point**

Hence, the class label for the new data point is 1.

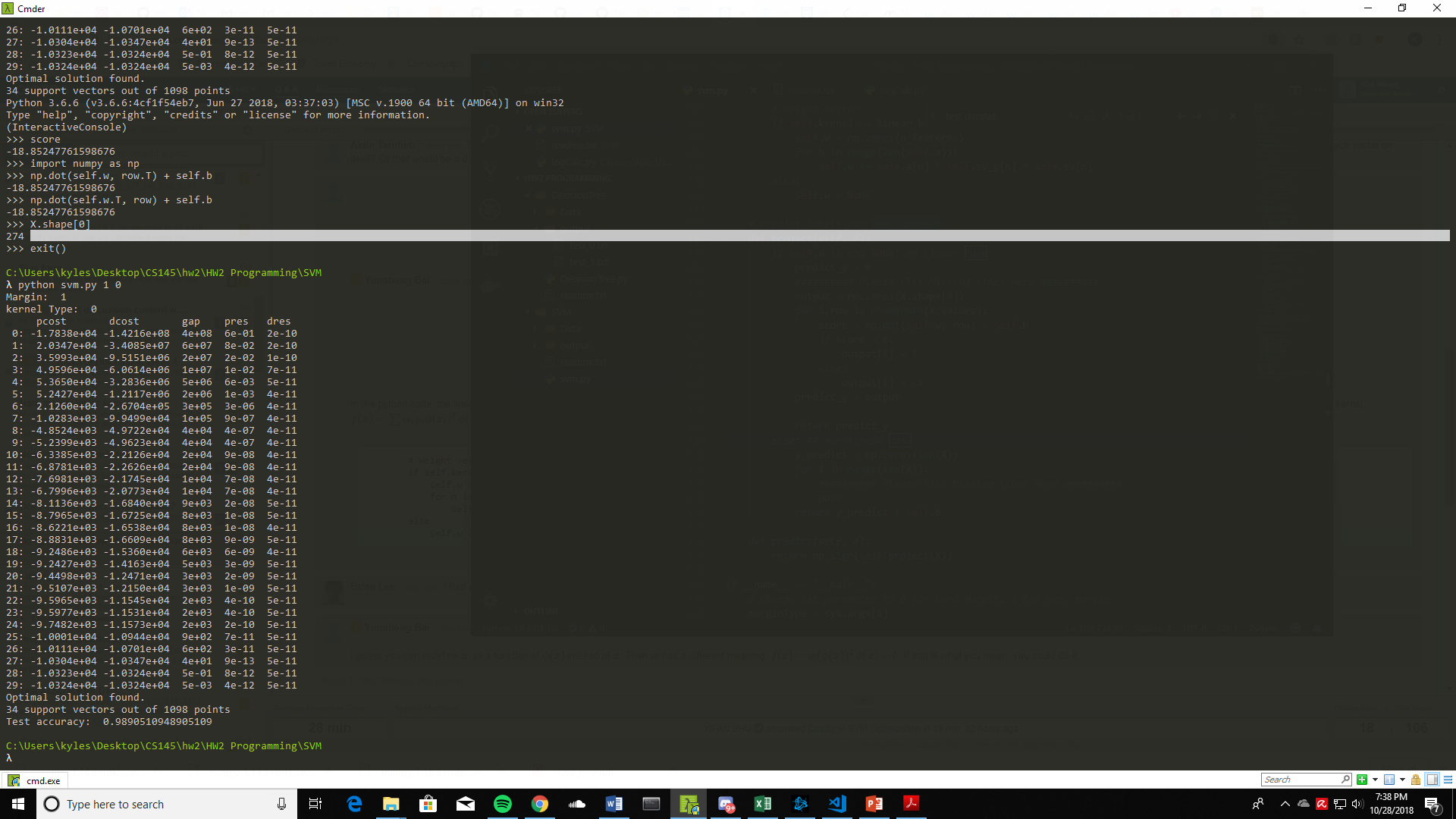
**2.1(f):** **Plot data points and decision boundary line**

Green data points are the support vectors.

**2.2(a):** **Hard margin and soft margin SVM for linear classifier**

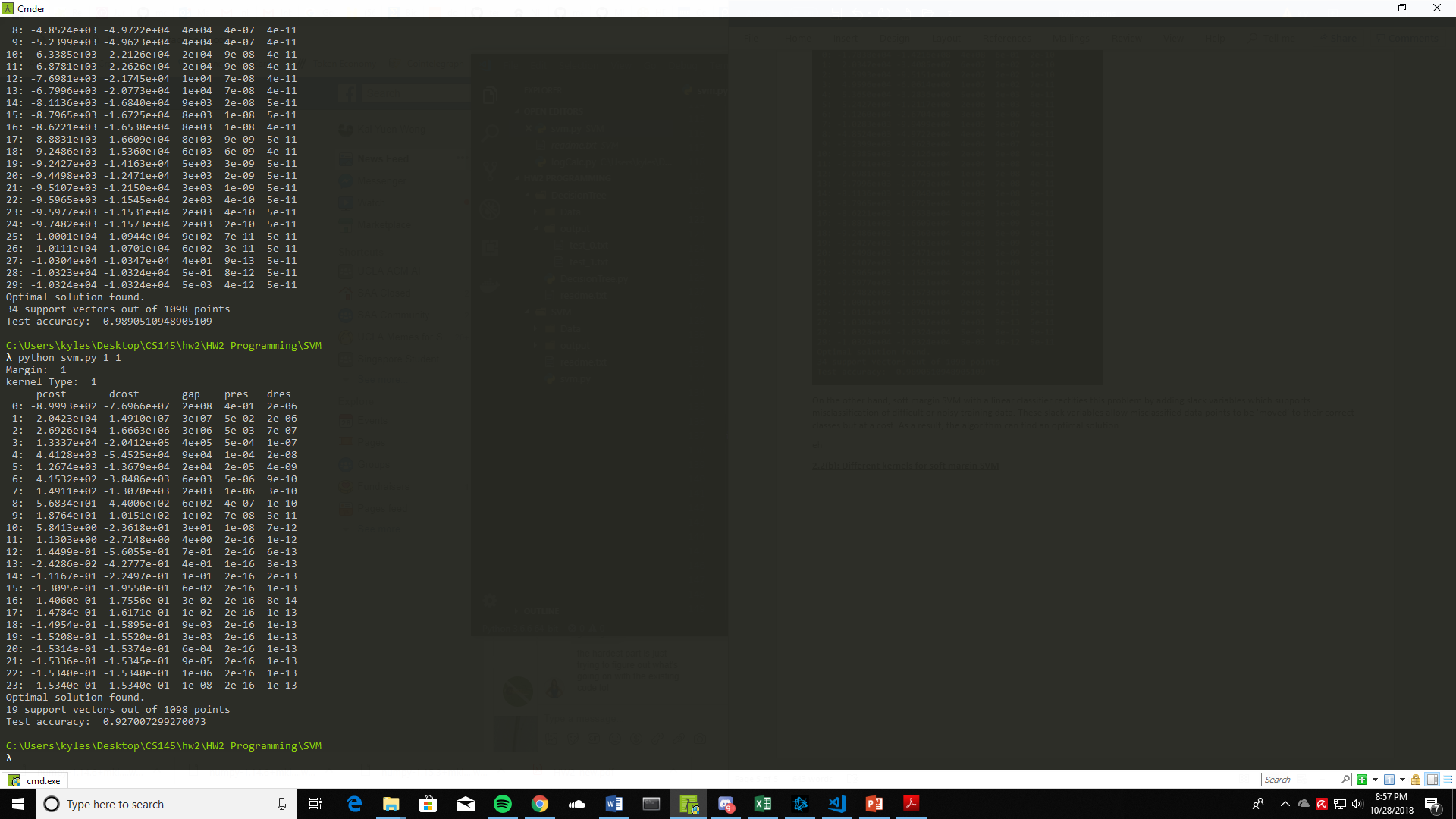


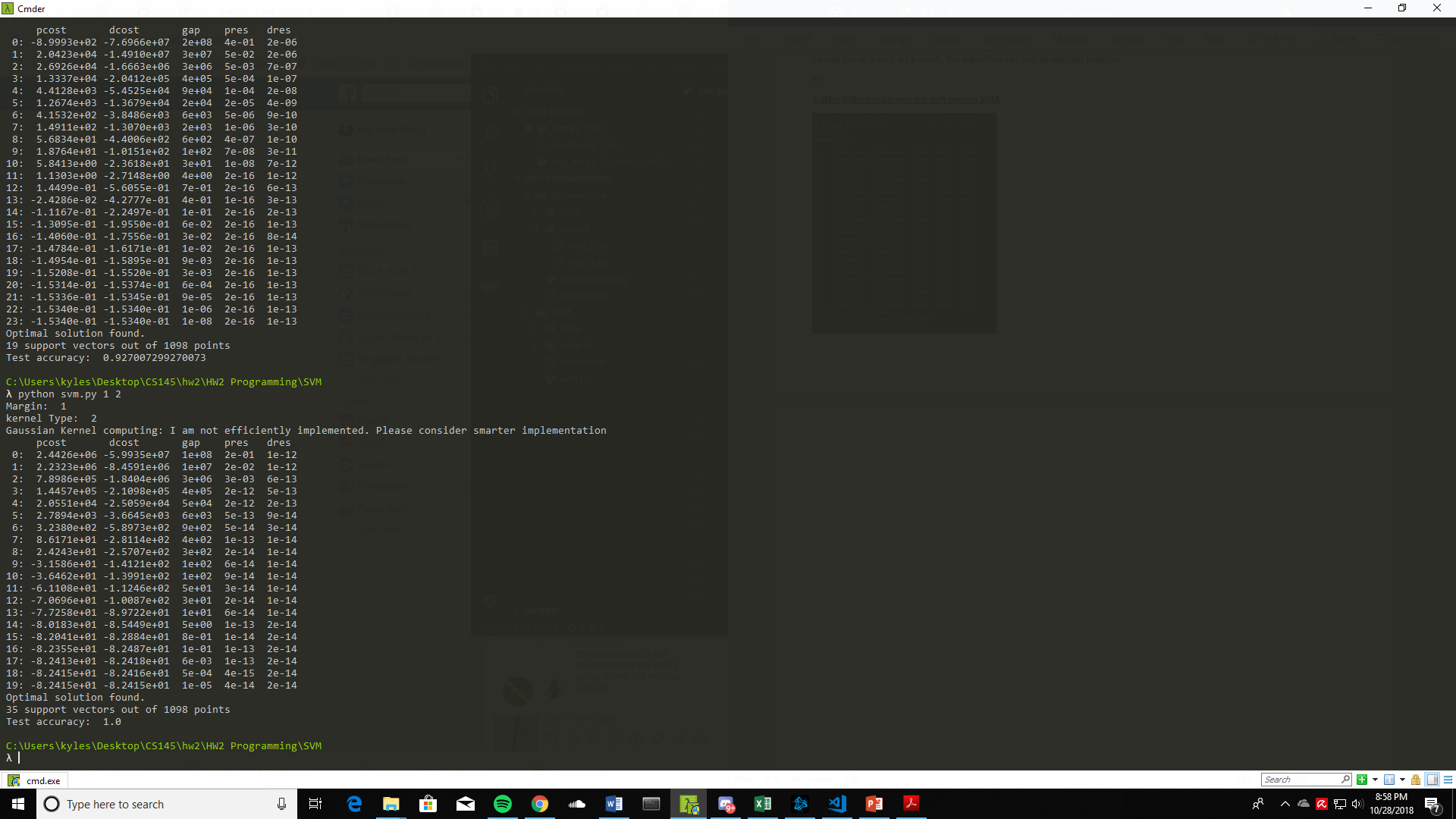
Hard margin SVM with a linear classifier terminates before finding an optimal solution because the training data is not linearly separable. As a result, a singular KKT matrix is generated by the cvxopt quadratic programming solvers which cannot be used to find the Lagrange multipliers.



On the other hand, soft margin SVM with a linear classifier rectifies this problem by adding slack variables which supports misclassification of difficult or noisy training data. These slack variables allow misclassified data points to be ‘moved’ to their correct classes, but at a cost. As a result, the algorithm can find an optimal solution.

**2.2(b):** **Different kernels for soft margin SVM**





I have the following accuracy results:

Soft margin SVM with linear kernel: 0.989

Soft margin SVM with polynomial kernel: 0.927

Soft margin SVM with gaussian kernel: 1.000

Based on this data, it appears that using the gaussian kernel gives the best model for this dataset and is thus the model I would want to choose. However, 100% accuracy is practically unrealistic and may indicate overfitting or insufficient size/complexity of the training dataset. Nonetheless, for the purposes of this assignment soft margin SVM with gaussian kernel is the best model for this dataset.