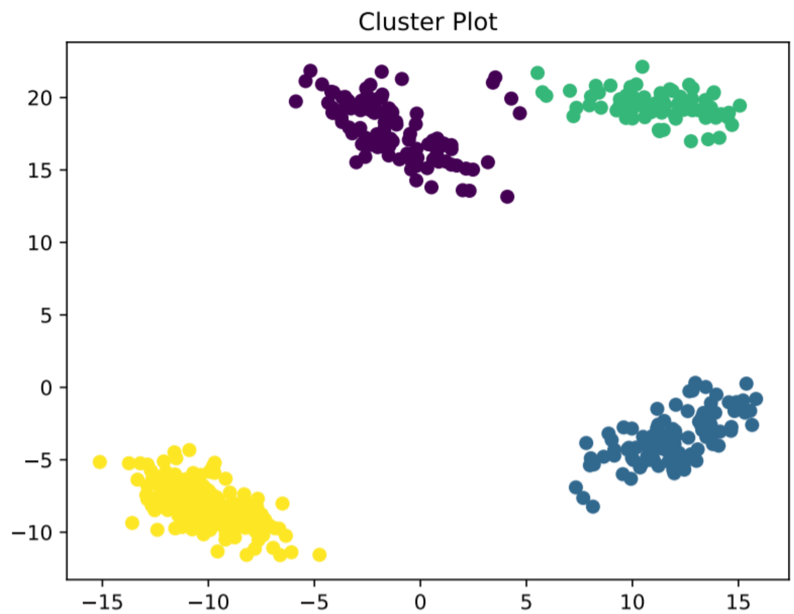
1 K-Means Clustering

1.1 Selecting among initializations

1. Code: https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/code/kmeans.py

2. Code: https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/code/main.py
   Line 32



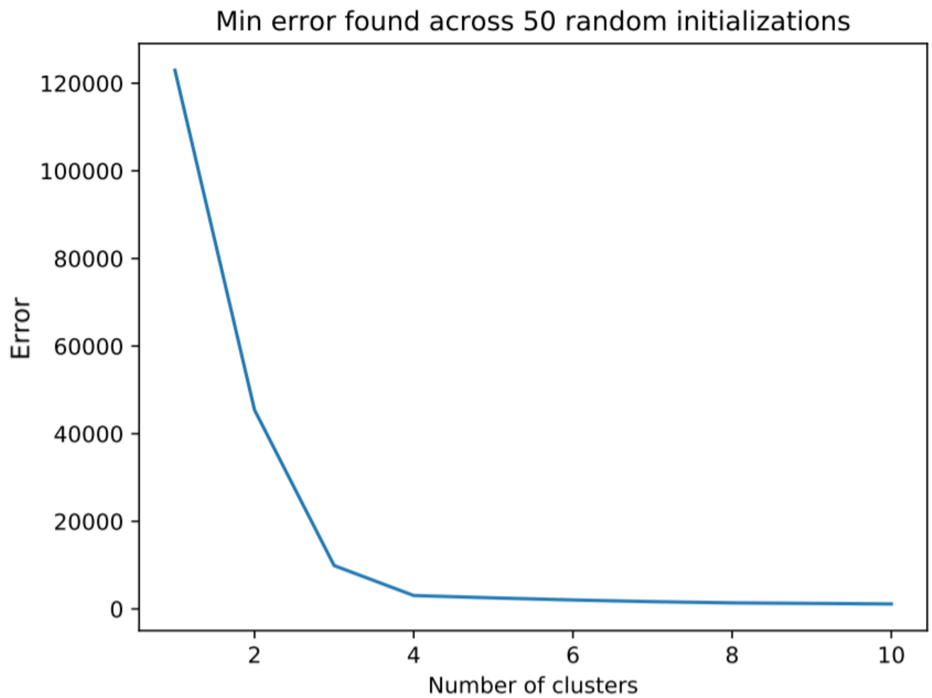
https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/figs/1.1_kmeans_lowest_err.pdf

1.2 Selecting k

1. As number of clusters or k increases, the distance between the points in the cluster and the cluster mean decreases. In this case, the value of the objective function will be the lowest if the largest possible value of k is chosen. Choosing the largest value of k will not give a good prediction of clusters.

2. Even with test data, the distance between the points in the cluster and the cluster mean will be the smallest with biggest k. Hence, the biggest k still get chosen.

3. Code: https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/code/main.py
   Line 55

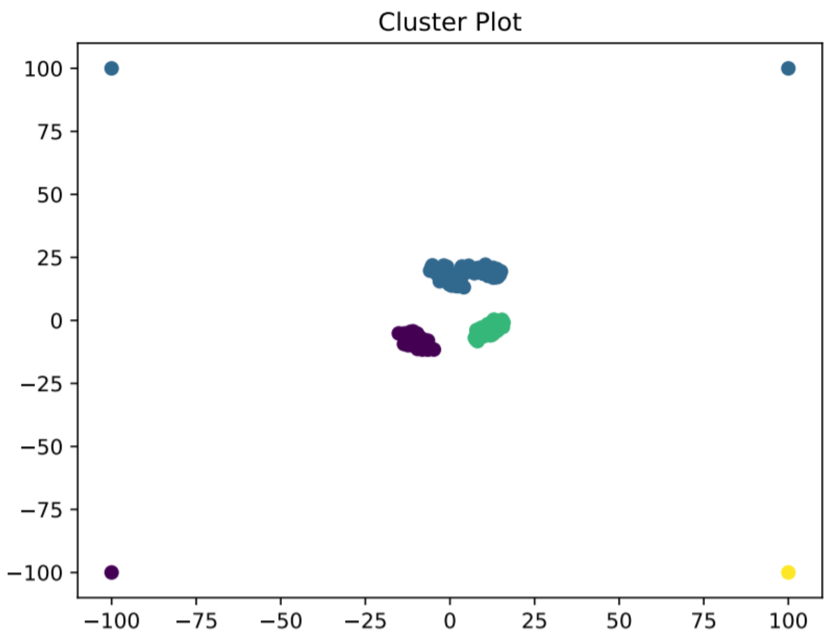


https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/figs/1.2_kmeans_err_with_k.pdf

4. According to the elbow method, the significant changes in slope from k: 1 -> 2, 2 -> 3, and 3 -> 4 would give reasonable values of k = 2, 3 and 4.

1.3 k Medians

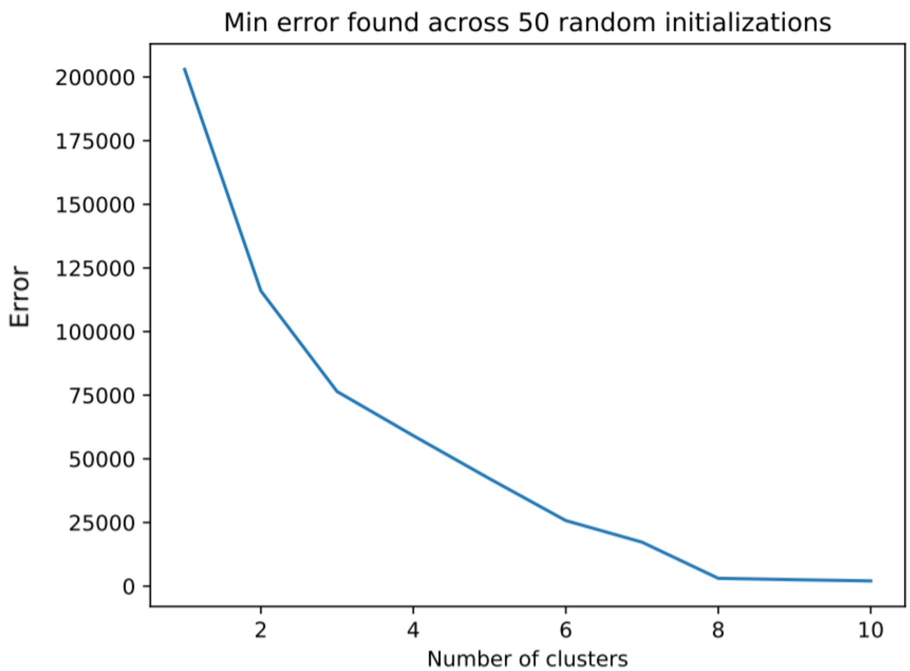
1. Code: https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/code/main.py
   Line 88



https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/figs/1.3.1_kmeans_outliers.pdf

The resulting prediction is not satisfying. For instance, the yellow outlier at the bottom right should not be a cluster by itself.

2. Code: https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/code/main.py
   Line 109

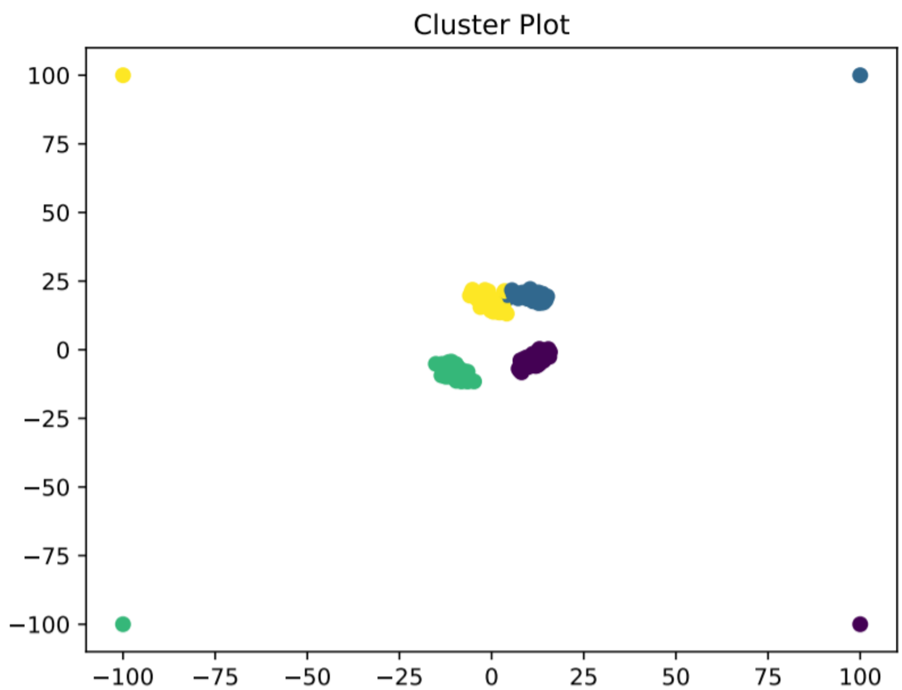


   https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/figs/1.3.2_kmeans_err_with_outlier.pdf

   According to the elbow method, the significant changes in slope from k: 1 -> 2, 2 -> 3, and 3 -> 8 would give reasonable values of k = 2, 3 and 8.

3. Code: https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/code/kmedians.py

4. Code: https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/code/main.py
Line 138

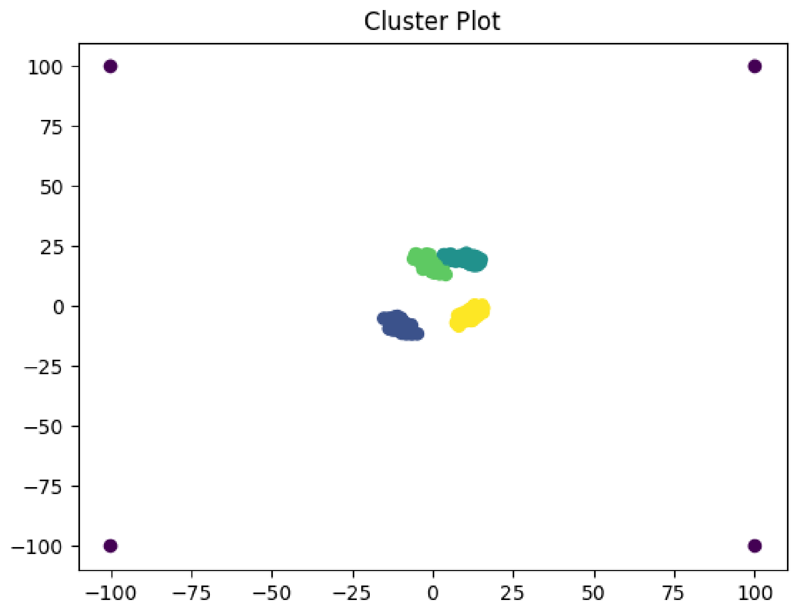


https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/figs/1.3.4_kmedians_outliers.pdf

The result is satisfying. There are now 4 clusters in the middle and the outliers are not a cluster by itself (eg. yellow point from part 1.3.1).

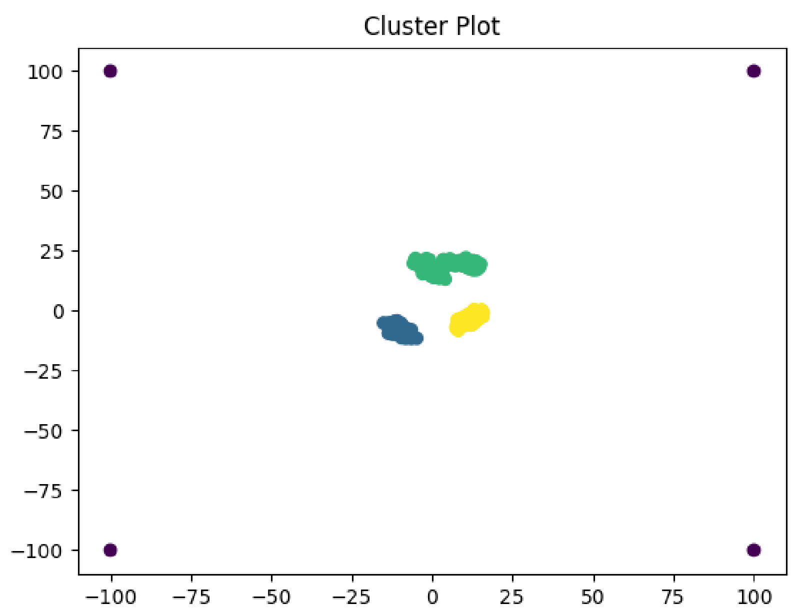
1.4 The Effect of Parameters on DBSCAN

1. min_samples = 3 and eps = 2



https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/figs/1.4.3_clusterdata_4_dbscan.pdf

2. min_samples = 5 and eps = 5



https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/figs/1.4.3_clusterdata_3_dbscan.pdf

3. min_samples = 5 and eps = 15

https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/figs/1.4.3_clusterdata_2_dbscan.pdf

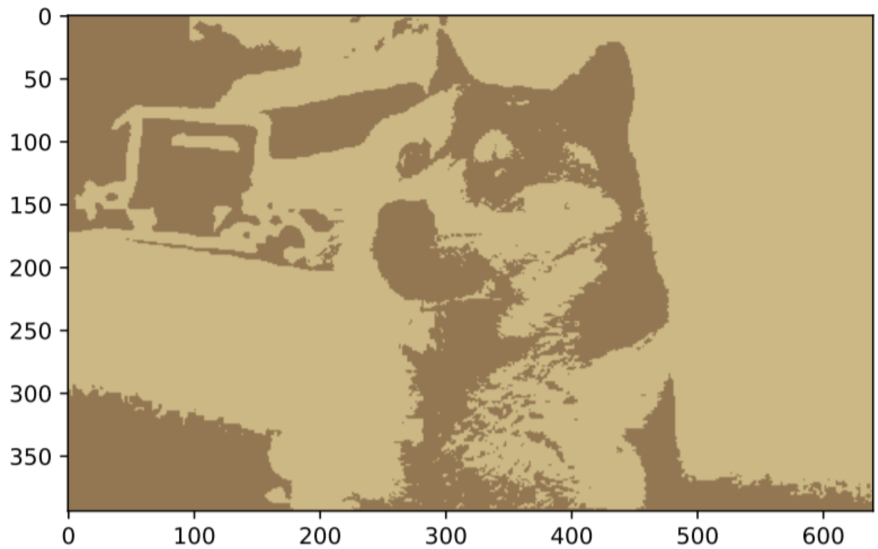
4. min_samples = 5 and eps = 20

https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/figs/1.4.3_clusterdata_1_dbscan.pdf

2 Vector Quantization and Density-Based Clustering

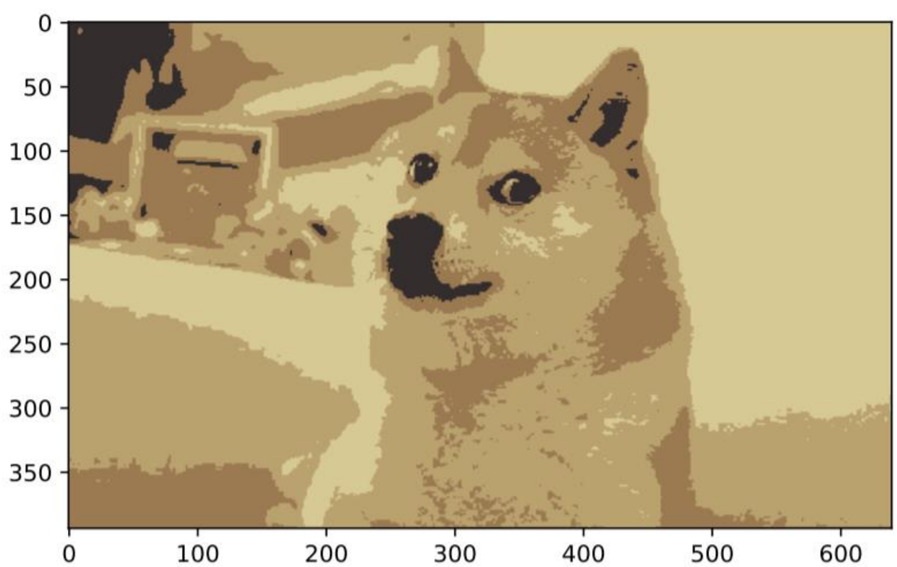
Code: https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/code/quantize_image.py

b = 1

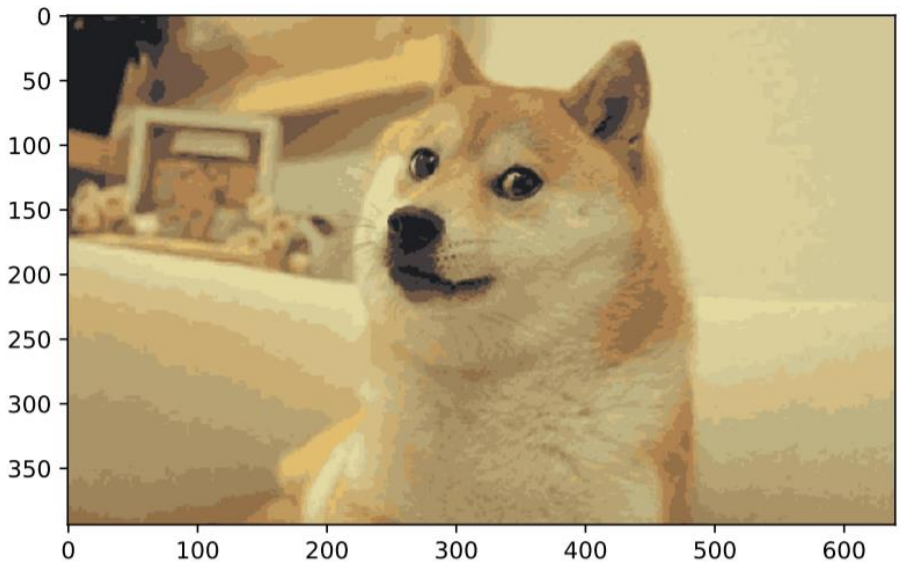


https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/figs/2_quantized_dog_1.pdf

b = 2



https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/figs/2_quantized_dog_2.pdf b = 4



https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/figs/2_quantized_dog_4.pdf

b = 6

https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/figs/2_quantized_dog_6.pdf

3 Vectors, Matrices and Quadratic Functions

3.1

1. $x^T x = \begin{bmatrix} 2 & -3 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \end{bmatrix} = 2 * 2 + (-3)(-3) = 13$

2. $\|x\|^2 = x^T x = 13$

3. $x^T(x + \alpha y) = x^T x + \alpha x^T y = 13 + 5 * (2 * 1 + (-3) * 4) = 13 - 50 = -37$

4. $Ax = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & -2 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \end{bmatrix} = \begin{bmatrix} 2-6 \\ 4-9 \\ 6+6 \end{bmatrix} = \begin{bmatrix} -4 \\ -5 \\ 12 \end{bmatrix}$

5. $z^T A x = \begin{bmatrix} 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} -4 \\ -5 \\ 12 \end{bmatrix} = 2 * (-4) + 12 = 4$

6. $A^T A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & -2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & -2 \end{bmatrix} = \begin{bmatrix} 1+4+9 & 2+6-6 \\ 2+6-6 & 4+9+4 \end{bmatrix} = \begin{bmatrix} 14 & 2 \\ 2 & 17 \end{bmatrix}$

7. True, $y\|y\|^2 = \|y\|^2 y$

8. True, $x^T A^T A z = z^T A^T A x$

9. False, $LHS: (1 * d), RHS: (d * 1)$

10. True, $(BC)^T = C^T B^T$

11. False, $-2x^T y \neq -x^T y$

12. True, $-y^T x + x^T y = 0$

3.2

1. $\sum_{i=1}^{n} |w^T x_i - y_i| = \|Xw - y\|_1$

2. $\|\mathrm{Xw} - \mathrm{y}\|_\infty + \frac{\lambda}{2}\|w\|_2^2$

3.3

1. $f(w) = \frac{1}{2}\|w - v\|^2 = \frac{1}{2}\|w\|^2 - w^T v + \frac{1}{2}\|v\|^2$

   $$\nabla f(w) = \frac{1}{2} * 2 * w - v = w - v$$

   $Setitng\ \nabla f(w) = 0\ gives\ w = v$

2. $f(w) = \frac{1}{2}\|w\|^2 + w^T X^T y$

   $$\nabla f(w) = \frac{1}{2} * 2 * w + X^T y = w + X^T y$$

   $Setting\ \nabla f(w) = 0\ gives\ w = -X^T y$

3. $f(w) = \frac{1}{2}\sum_{i=1}^{n} z_i\,(w^T x_i - y_i)^2 = \frac{1}{2} Z\|Xw - y\|^2 = \frac{1}{2}(Xw - y)^T Z(Xw - y)$

   $$= \frac{1}{2}(w^T X^T Z X w - w^T X^T Z y - y^T Z X w + y^T Z y)$$

   $$= \frac{1}{2}(w^T X^T Z X w - 2w^T X^T Z y + y^T Z y)$$

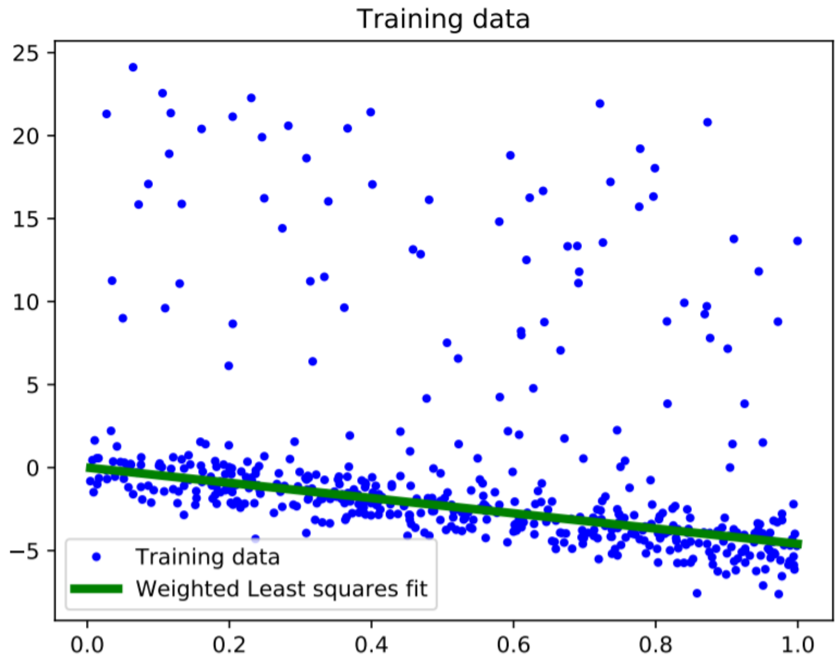
   $$= \frac{1}{2} w^T X^T Z X w - w^T X^T Z y + \frac{1}{2} y^T Z y$$

   $\nabla f(w) = X^T Z X w - X^T Z y$

   $Setting\ \nabla f(w) = 0\ and\ solve\ X^T Z X w = X^T Z y$

4 Vectors, Matrices and Quadratic Functions

4.1 Code: https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/code/linear_model.py



https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/figs/weighted_least_squares_outliers.pdf

4.2

$$f(w) = \sum_{i=1}^{n} \log((e^{w^T x_i - y_i}) + (e^{y_i - w^T x_i}))$$

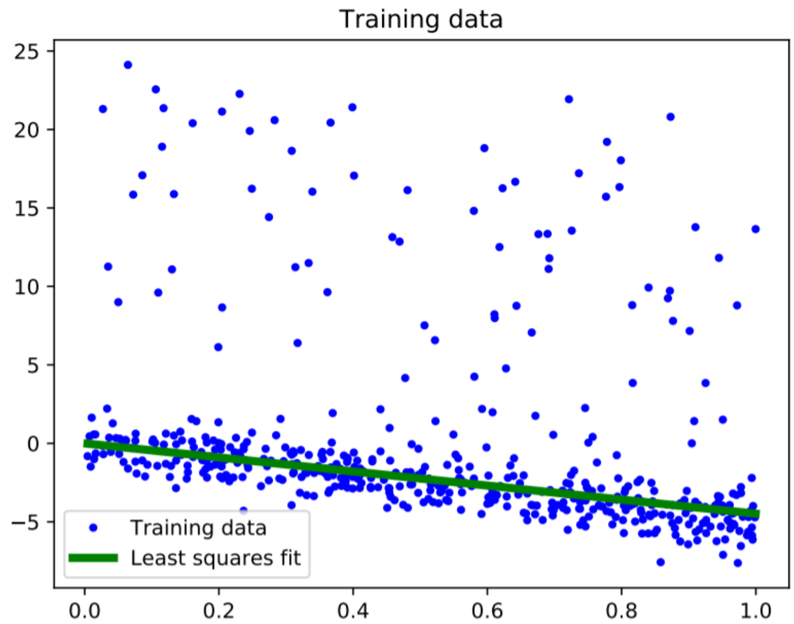
let $r_i = (w^T x_i - y_i)$, so $f(w) = \sum_{i=1}^{n} \log((e^{r_i}) + (e^{-r_i}))$

$$\text{Since } \nabla f(w) = \frac{\partial f(w)}{\partial w_i} = \sum_{i=1}^{n} \frac{e^r - e^{-r}}{e^r + e^{-r}} * x_{ij}$$

$$\frac{e^r - e^{-r}}{e^r + e^{-r}} \; is\ a\ scalar, and\ let\ r = \frac{e^r - e^{-r}}{e^r + e^{-r}},$$

$$so\ \nabla f(w) = \sum_{i=1}^{n} r_i * x_i = X^T r$$

4.3 Code: https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/code/linear_model.py



https://github.ubc.ca/cpsc340-2017S/sopida_zhenxil_a2/tree/master/figs/gradient_descent_model.pdf