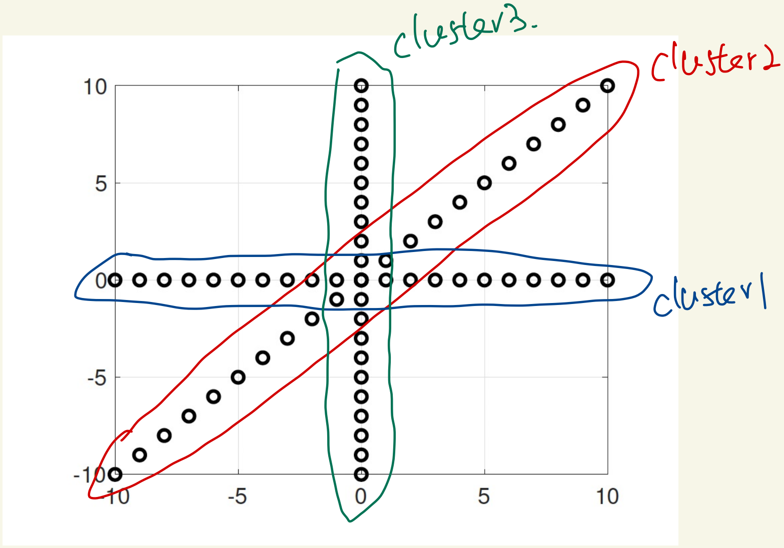
CS541_final_exam

Chenxu Wang

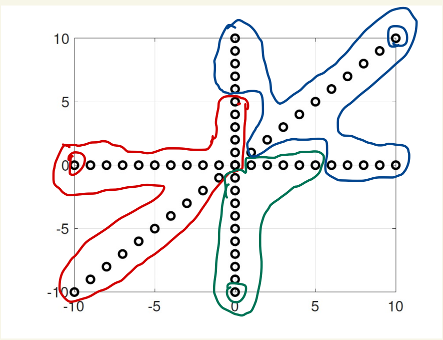
10457625

0. Chenxu Wang. 10457625

1.1



1.2



When initial centers are (-10, 0), (0, -10), (10, 10)

K=3.

We iterates the points from left to right. and assign points in the order (-10, 0), (0, -10), and (10, 10).

2.

Table 1: Patient data.

	Age	Weight	Height	Gender	Blood Pressure	$\cdots$	Sharp Pain
Patient 1	z_{11}	z_{12}	z_{13}	z_{14}	?	$\cdots$	z_{1m}
Patient 2	z_{21}	z_{22}	z_{23}	z_{24}	z_{25}	$\cdots$	z_{2m}
Patient 3	z_{31}	z_{32}	z_{33}	z_{34}	?	$\cdots$	z_{3m}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Patient n	z_{n1}	z_{n2}	z_{n3}	z_{n4}	z_{n5}	$\cdots$	z_{nm}

We have $m-1$ number of feasures. and the prediction label is Blood Pressure. and n samples.

First step, we have to normalize the data we choose to use max-min normalization.

Secondly, we need to reduce the dimension. we can use PCA to find some principle components.

Finally, we can apply linear regression to predict the Blood Presure.

3.

Table 2: Patient data.

	Age	Weight	Height	Gender	Blood Pressure	$\cdots$	Sharp Pain
Patient 1	?	z_{12}	z_{13}	z_{14}	?	$\cdots$	z_{1m}
Patient 2	?	z_{22}	?	z_{24}	z_{25}	$\cdots$	?
Patient 3	z_{31}	?	z_{33}	?	?	$\cdots$	z_{3m}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Patient n	?	z_{n2}	z_{n3}	?	z_{n5}	$\cdots$	?

Since we have a lot of missing values, we can build the sparse matrix $\mathbf{X}$. Let all of "?" be 0.

x_1 represents Patient 1. ($\mathbb{R}^m$) which is a column vector.

So, we can have $\{x_1 \cdots x_n\} \in \mathbb{R}^m$, next normalise data.

Then, we can apply dictionary learning for estimating all the missing values.

$\therefore$ Dictionary: $Z = \begin{pmatrix} \uparrow & & \uparrow \\ Z_1 & \cdots & Z_r \\ \downarrow & & \downarrow \end{pmatrix} \in \mathbb{R}^{m \times r}$

weights: $w_i \in \mathbb{R}^r$

Next we can have: $\min\limits_{Z, w_1 \cdots w_n} \frac{1}{n} \sum_{i=1}^{n} \left(\| x_i - Z w_i \|_2^2 + \lambda \| w_i \|_1 \right)$

s.t. $\| Z_j \|_2 \leq 1$, $j = 1 \cdots r$

finally, we can apply Alternating Minimization to calculate $Z, w_1 \cdots w_n$. After getting Z_{opt}, W_{opt}, we get prediction matrix $Z_{opt} W_{opt}$.

4.

For preventing false positive, we have to find the really right samples with a high probability to be false.

We can apply logistic regression to build the model and obtain the "w".

Next, we think above PAC learning.

To assume distribution D over $\mathbb{R}^d \times \{-1, +1\}$. all (x, y) from D. Given "w" above, we have.

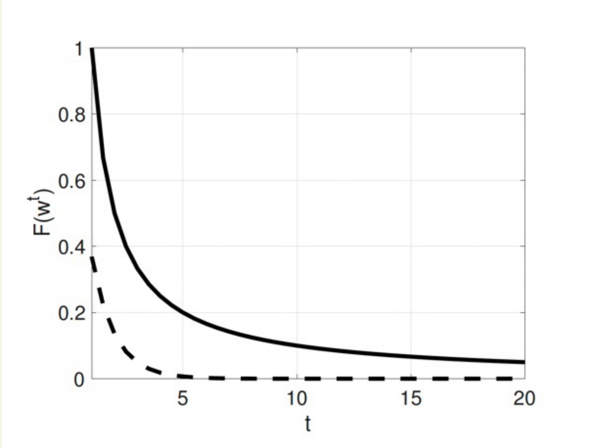
$err_D(w) = Pr_{(x,y)\in D}(y = sign(w \cdot x))$., (x,y) randomly drawn from D.

The let $n = \frac{1}{\varepsilon} \cdot (d + \log\frac{1}{\delta})$, draw the training data $(x_1, y_1) \cdots (x_n, y_n) \sim D$., then apply SVM to obtain w., with probability $1-\delta$, $err_D(w) \leq \varepsilon$.

For our problem, we have to set δ enough small, like 0.01. to guarantee the correction rate.

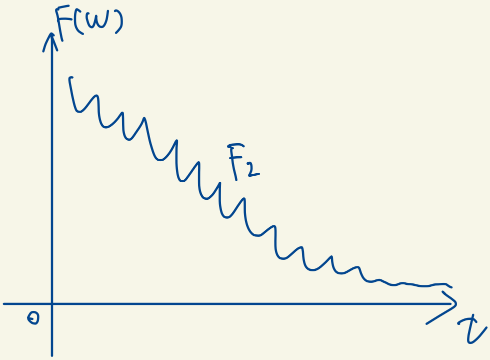
Another way, boosting may can help too, because it's an approach to reduce misclassification error.

5



①. The dashed line may correspond to F_1 (smooth), because smooth function converge faster.

②.



We have F_2 is strongly convex and non-smooth one. Therefore. $\eta_t = O(\frac{1}{t})$.

F(w) may be fluctuate when going gradient descent. and it will converge.