學號：B10705025： 資管三　姓名：彭鈞道

這次的python檔也包含下面用DNN做feature transformation，該部分要跑較久，SVM的地方應該很快。

1. (2%) 請說明你是如何normalize discrete跟continous的feature

Continous的部分使用一般的normalization，各項減去平均值再除以標準差。Discrete的部分如果用kaggle上面的資料集，應該用one hot vector來表示各項參數的indicator，但助教的sample code已經處理好這些問題，所以便沒有額外多做。

2. (2%) 使用DNN做 feature transformation，將output dimension設為4跟1024，並將其output丟進linear SVM訓練，比較leaderboard上的結果，並說明造成這樣結果的原因

（hint：linear SVM本身是linear classifier，資料必須是linearly separable的資料）

1024的效果比4好，因為在越多維的空間中，越容易找到linear separate的hyperplane，所以1024維有更好的linear separate效果。

Submission and Description	Private Score	Public Score	Selected
predict for output 1024 (1).csv Complete · 4d ago	0.83613	0.83931	☐
predict for output 4 (1).csv Complete · 4d ago	0.78319	0.77874	☐

3. (6+2%) Refer to math problem

https://ntueemlta2023.github.io/homeworks/hw5/ml-2023fall-hw5-math.pdf

Math 1

using polynomial kernal $k(x,y) = (\langle x,y \rangle + c)^d$

hw 5. (1) choose $c=0, d=2$

$$k(x,y) = \langle \varphi(x), \varphi(y) \rangle = (x^T y)^2$$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$(x^T y)^2 = (x_1 y_1 + x_2 y_2)^2$$

$$= ((x_1 y_1)^2 + (x_2 y_2)^2 + 2 x_1 y_1 x_2 y_2)$$

$$= \langle \varphi(x), \varphi(y) \rangle$$

$$\varphi(x) = \begin{bmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2} x_1 x_2 \end{bmatrix}, \quad \varphi(y) = \begin{bmatrix} y_1^2 \\ y_2^2 \\ \sqrt{2} y_1 y_2 \end{bmatrix}$$

feature map $\varphi(z) = \begin{bmatrix} z_1^2 \\ z_2^2 \\ \sqrt{2} z_1 z_2 \end{bmatrix}$

MATH2

hw5 $f(x)=\sum_{i=1}^{N}\alpha_i y_i k(x_i,x)+b$, $b=0$, $\alpha_i=1$

2. $=\sum_{i=1}^{N} y_i k(x_i,x)+b$

$|f(x_i)-y_i|<1$

$\Rightarrow \left|\sum_{j=1}^{N} y_j e^{\frac{-\|x_j-x_i\|^2}{T^2}} - y_i\right| < 1$

$= \left|-\sum_{j=1, y_j=-1}^{N} y_j e^{\frac{-\|x_j-x_i\|^2}{T^2}} + \sum_{j=1, y_j=1}^{N} y_j e^{\frac{-\|x_j-x_i\|^2}{T^2}} - y_i\right| < 1$

$\|x_j-x_i\|\geq\varepsilon$ $\|x_i-x_j\|^2\geq\varepsilon^2$ $e^{-\|x_j-x_i\|^2}\leq e^{-\varepsilon^2}$

let $L=-\sum_{j=1, y_j=-1}^{N} e^{\frac{-\|x_j-x_i\|^2}{T^2}} + \sum_{j=1, y_j=1}^{N} e^{\frac{-\|x_j-x_i\|^2}{T^2}}$

$L = -N_1 e^{\frac{-\|x_j-x_i\|^2}{T^2}} + N_2 e^{\frac{-\|x_j-x_i\|^2}{T^2}}$

number of classified as 1 number of classified as 2

$-N_1 e^{\frac{+\varepsilon^2}{T^2}} \leq L \leq N_2 \cdot e^{\frac{-\varepsilon^2}{T^2}}$

prove $|L-y_i|<1 \Rightarrow$ failed

try $\left|\sum_{j=1}^{N} y_j e^{\frac{-\|x_j-x_i\|^2}{T^2}} - y_i\right| < 1$

$= \left|\sum_{j=1, j\neq 1} y_j e^{\frac{-\|x_j-x_i\|^2}{T^2}} + y_i - y_i\right| < 1$

let $n_1=\sum_{j=1, j\neq i, y_j=1}^{N} y_j$ $\Rightarrow$ let $\underbrace{\left|\sum_{j=1, j\neq 1} y_j e^{\frac{-\|x_j-x_i\|^2}{T^2}}\right|}_{\text{equals } L-y_i} < 1$

$n_2=-\sum_{j, j\neq i, y_j=-1} y_j$ $\left|n_1 e^{\frac{\|x_j-x_i\|^2}{T^2}} - n_2 e^{\frac{-\|x_j-x_i\|^2}{T^2}}\right| < 1$

$$-n_2 e^{\frac{-\varepsilon^2}{T^2}} \le n_1 e^{\frac{-\|x_j - x_\ell\|^2}{T^2}} - n_2 e^{\frac{-\|x_j - x_i\|^2}{T^2}} \le n_1 e^{\frac{-\varepsilon^2}{T^2}}$$

prove $\Rightarrow n_2 e^{\frac{-\varepsilon^2}{T^2}} \ge 1, \quad \frac{-\varepsilon^2}{T^2} \le \log n_2$

$n_1 e^{\frac{-\varepsilon^2}{T^2}} \le 1, \quad \frac{-\varepsilon^2}{T^2} \le \log n_1 \quad \Rightarrow T \le \sqrt{\frac{\varepsilon^2}{\log n_2}}$

$n_1, n_2 \neq 0$

$, \quad \frac{\varepsilon^2}{T^2} \ge \log n_1 \quad \Rightarrow T \le \sqrt{\frac{\varepsilon^2}{\log n_2}}$

$$\Rightarrow T \le \min\left(\sqrt{\frac{\varepsilon^2}{\log\left(\sum_{j \mid j \neq i, y_j = 1} y_j\right)}}, \sqrt{\frac{\varepsilon^2}{\log\left(\sum_{j; j \neq i, y_j = 1} -y_i\right)}} \right)$$

if n_1 or $n_2 = 0$

$\Rightarrow$ simply find a line that make $f(x_i) = 1$. forall i

$= 1$, forall i

hw5

3.(a) rewrite the problem to $m\times n$ $\quad x_i = n\times 1$

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i$$

$$\text{s.t. } y_i - w^T x_i - b - \epsilon - \xi_i \le 0 \quad i=1,\dots,m$$

$$w^T x_i + b - y_i - \epsilon - \xi_i \le 0 \quad i=1,\dots,m$$

$$-\xi_i \le 0 \quad i=1,\dots,m$$

Lagrangian $(w,b,\xi,\alpha,\alpha^*,\beta)$

$$= \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{m}\xi_i + \sum_{i=1}^{m}\alpha_i(y_i - w^T x_i - b - \epsilon - \xi_i)$$

$$+ \sum_{i=1}^{m}\alpha_i^*(w^T x_i + b - y_i - \epsilon - \xi_i) + \sum_{i=1}^{m}\beta(-\xi_i)$$

3.(b) Maximize $\inf_{w,b\in D}\{\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i + \sum_{i=1}^{m}\alpha_i(y_i - w^T x_i - b - \epsilon - \xi_i)$

$$+\sum_{i=1}^{m}\alpha_i^*(w^T x_i + b - y_i - \epsilon - \xi_i) + \sum_{i=1}^{m}\beta_i(-\xi_i)\}$$

let $\xi = \begin{bmatrix}\xi_i\\ \vdots \\ \xi_m\end{bmatrix}$

subject to $\alpha_i \ge 0, \ i=1,\dots,m$

$\alpha_i^* \ge 0, \ i=1,\dots,m$

$\beta_i \ge 0, \ i=1,\dots,m$

$(\beta+\Delta\beta)^T(-\xi)$

$= \beta\cdot(-\xi) + (\Delta\beta)^T\cdot(-\xi)$

整理 target function

$1\times 2 \times 2\times 3\times 3\times m\times n+1$

$$\frac{1}{2}w^T w + C\cdot\xi + \alpha^T(y - w^T X - b\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} - \epsilon\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} - \xi)$$

$$+ \alpha^{*T}(w^T X + b\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} - y - \epsilon\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} - \xi)$$

$$+ \beta^T(-\xi) = \theta$$

take $\frac{\partial L}{\partial b} = -\alpha^T\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} + \alpha^{*T}\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} = (\alpha^{*T} - \alpha^T)\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix}$

$$= \sum_{i=1}^{m}(\alpha^*_i - \alpha_i)$$

$$\frac{\partial L}{\partial W} = W - \sum_{i=1}^{m}\alpha_i X_i + \sum_{i=1}^{m}(\alpha^* X_i) = W + \sum_{i=1}^{m}(\alpha^*_i - \alpha_i)\cdot X$$

$$\frac{\partial L}{\partial \xi} = C\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} - \alpha - \alpha^* - \beta \Rightarrow \frac{\partial L}{\partial \xi_i} = C - \alpha_i - \alpha^*_i - \beta_i$$

3.(c) $\overline{w} = \sum_{i=1}^{m} (\alpha_i - \overline{\alpha_i^*}) x_i$ (by $\frac{\partial \theta}{\partial w} = w + \sum_{i=1}^{m} (\alpha^*_i - \alpha) \underset{=0}{x_i}$)

(1) by the inequality $y_i - w^T x_i - b \le \epsilon + \xi_i$

$w^T x_i + b - y_i \le \epsilon + \xi_i$

$\xi_i \ge 0$

$\Rightarrow |y_i - w^T x_i - b| \le \epsilon + \xi_i$

$\Rightarrow \xi_i \ge |y_i - w^T x_i - b| - \epsilon$, and $\xi_i \ge 0$

$\Rightarrow \xi_i \ge \max(0, |y_i - w^T x_i - b| - \epsilon)$

the objective becomes

minimize $\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \max(|y_i - w^T x_i - b| - \epsilon, 0)$

So $\overline{w}, \overline{b} = \underset{w, b \in R}{\text{argmin}}$ objective function

$= \underset{b \in R}{\text{argmin}} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \max(|y_i - (\overline{w}^T x_i + b)| - \epsilon, 0)$

but $\overline{w}$ are already know

$\Rightarrow \overline{b} = \underset{b \in R}{\text{argmin}}\ C\sum_{i=1}^{m} \max(|y_i - (w^T x_i + b)| - \epsilon, 0)$ ✱

(2) The primal problem holds slater's condition

domain is a convex set, objective, constraints are convex functions, and we can find strictly feasible point (set ϵ large)

apply KKT, stationary conditions

for $(\overline{w}, \overline{b}, \overline{\xi}, \overline{\alpha}, \overline{\alpha^*}, \overline{\beta})$

KKT

Feasibility

$y_i - \overline{w}^T x_i - \overline{b} \le \epsilon + \overline{\xi}_i$

$\overline{w}^T x_i + \overline{b} - y_i \le \epsilon + \overline{\xi}_i$

$\overline{\xi}_i \ge 0$

$\overline{\alpha}_i \ge 0, \overline{\alpha}^*_i \ge 0, \overline{\beta}_i \ge 0$

$\overline{\beta}_i \overline{\xi}_i = 0, \overline{\alpha}_i (y_i - \overline{w}^T x_i - \overline{b} - \epsilon - \overline{\xi}_i) = 0, \overline{\alpha}^*_i (\overline{w}^T x_i + \overline{b} - y_i - \epsilon - \overline{\xi}_i)$

derivatives $=0$ in 3.(b)

$\sum_{i=1}^{M} \overline{\alpha}^*_i - \overline{\alpha}_i = 0$

$C - \overline{\alpha}_i - \overline{\alpha}^*_i - \overline{\beta}_i = 0$

$w = \sum_{i=1}^{m} (\overline{\alpha}_i - \alpha^*_i) x_i$

3.(c) by the conditions above

(2) $e = y_i - \bar{w}^T x_i - b$,

$|e| < \varepsilon$, $(e - \epsilon - \bar{\xi}_i)$ and $(-e - \epsilon - \xi_i) \neq 0$

$\Rightarrow \bar{\alpha}_i = 0$, and $\bar{\alpha}^*_i = 0$ #

$\Rightarrow \xi_i$ in objective, it should be minimal, and $\bar{\beta}_i \bar{\xi}_i = 0$, $\beta_i \geq 0$

$\Rightarrow \bar{\xi}_i = 0$ #

$e = \epsilon$, $\bar{\alpha}_i (-\xi_i) = 0$, $\bar{\alpha}^*_i (-2\epsilon - \xi_i) = 0$

≤ 0 (by feasibility) ≤ 0 (by feasibility)

$-2\epsilon - \xi_i \neq 0 \Rightarrow \bar{\alpha}^*_i = 0 \Rightarrow C - \bar{\alpha}_i - \underbrace{\bar{\beta}_i}_{\geq 0} = 0$

$\Rightarrow \bar{\xi}_i = 0$ (minimize), $0 \leq \bar{\alpha}_i \leq C$

$e = -\epsilon$, $\bar{\alpha}_i (-2\epsilon - \bar{\xi}_i) = 0$, $\bar{\alpha}^*_i (-\bar{\xi}_i) = 0$

$\Rightarrow \bar{\alpha}_i \neq 0$, $\bar{\xi}_i = 0$ (minimize) $\Rightarrow C - \underbrace{\bar{\alpha}^*_i}_{\geq 0} - \underbrace{\bar{\beta}_i}_{\geq 0} = 0$

$\Rightarrow 0 \leq \bar{\alpha}^*_i \leq C$

$e > \epsilon \Rightarrow e \leq \epsilon + \bar{\xi}_i$, $-e \leq \epsilon + \bar{\xi}_i \Rightarrow \bar{\xi}_i \geq e - \epsilon \geq 0$

$\Rightarrow \beta_i = 0$ $(\beta_i \xi_i = 0)$, let $\bar{\xi}_i = e - \epsilon$ (minimize)

$\Rightarrow \bar{\alpha}_i = 0$, $C - \underbrace{\bar{\alpha}_i}_{=0} - \underbrace{\bar{\beta}_i}_{=0} - \bar{\alpha}^*_i = 0$, $\bar{\alpha}^*_i = C$

$e < -\varepsilon$, $\Rightarrow$ $e \leq \epsilon + \bar{\xi}_i$, $-e \leq \epsilon + \bar{\xi}_i \Rightarrow \xi_i \geq -e - \varepsilon \geq 0$

$\Rightarrow \beta_i = 0$, let $\bar{\xi}_i = -e - \varepsilon \Rightarrow \bar{\alpha}_i = 0$

$\Rightarrow C - \bar{\alpha}_i - \bar{\beta}_i - \bar{\alpha}^*_i = 0 \Rightarrow \bar{\alpha}^*_i = C$

3.(d) by the derivatives in 3(b) get

$C = \alpha_i + \bar{\alpha}^*_i + \beta_i$, $\sum_{i=1}^{m} \alpha_i - \alpha^*_i = 0$, $W = \sum_{i=1}^{m} (\alpha_i - \alpha^*_i) x_i$

(1) $\theta = \inf \Big(\frac{1}{2}\|w\|^2 + \sum_{i=1}^{m} \alpha_i (y_i - w^T x_i - b - \epsilon - \xi_i) + \sum_{i=1}^{m} \alpha^*_i (w^T x_i - y_i + b - \epsilon - \xi_i) + \sum_{i=1}^{m} \beta_i (-\xi_i) + C \sum_{i=1}^{m} \xi_i \Big)$

$= \frac{1}{2}\|w\|^2 + \epsilon \sum_{i=1}^{m} (\alpha_i + \alpha^*_i) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\alpha_i - \alpha^*_i)(\alpha_j - \alpha^*_j) \underbrace{x_i^T x_j}_{\langle x_i, x_j \rangle}$

(2) $f(x) = \sum_{i=1}^{M} (\alpha_i - \alpha^*_i) \phi(x_i)^T \phi(x) - \epsilon \sum_{i=1}^{M} (\alpha_i + \alpha^*_i) + b$

learning $\max \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\alpha_i - \alpha^*_i)(\alpha_j - \alpha^*_j) \langle \phi(x_i)^T, \phi(x_j) \rangle$

Math4

hw5. 4.1

$$L(\alpha, b, \xi, w, \beta, \gamma)$$

$$= \frac{1}{2}\sum_{i=1}^{N}\alpha_i^2 + \sum_{i=1}^{N} c_i \xi_i + \sum_{i=1}^{N} w_i \left(1-\xi_i - y_i\left(\sum_{j=1}^{N}\alpha_j y_j x_i \cdot x_j + b\right)\right)$$

$$+ \sum_{i=1}^{N}\beta_i(-\alpha_i) + \sum_{i=1}^{N}\gamma_i(-\xi_i)$$

4.2 check slater's condition.

domain of variables $\alpha \in R_+^n$, $b \in R$, $\xi \in R_+^n$

$\Rightarrow$ all convex sets $\Rightarrow$ so the domain ~~are~~ is convex

$\frac{1}{2}\sum_{i=1}^{N}\alpha_i^2 = \frac{1}{2}\|\alpha\|_2^2 \Rightarrow$ 2-norm is convex, $\sum_{i=1}^{N} c_i \xi_i$ is convex

$\Rightarrow$ objective is convex, $-\alpha_i$ is convex, $-\xi_i$ is convex

$\underbrace{1-\xi_i}_{\text{linear}} - y_i \Big(\underbrace{\sum_{j=1}^{N}\alpha_i y_i x_i \cdot x_j + b}_{\text{linear combination } \alpha\text{'s elemenet}}\Big)$ is convex, no equality constraint, so $0 \in \mathrm{int}(h(x))$

, set ξ_i very large, $\alpha > 0$, we can simply find strictly feasible solution

$\Rightarrow$ by strong convex theorem $\Rightarrow$ duality gap $= 0$

4.3 $\frac{\partial L}{\partial \alpha_i} = \alpha_i + \sum_{j=1}^{N} w_j (-y_i)(y_j) \cancel{x_j} \cdot x_i + \beta_i$

$\frac{\partial L}{\partial b} = \sum_{i=1}^{N} w_i (-y_i) \implies \sum_{i=1}^{N} w_i(-y_i) = 0$

$\frac{\partial L}{\partial \varepsilon_i} = C_i - w_i - \gamma_i \gamma_i$

$$\theta(w,\beta,\gamma_i) = \inf_{\alpha, b, \varepsilon} \left(f(\alpha,b,\varepsilon) + \sum_{i=1}^{N} w_i (g_{1i}(\alpha,b,\varepsilon)) + \sum_{i=1}^{N} \beta_i (g_{2i}(\alpha,b,\varepsilon)) + \sum_{i=1}^{N} \gamma_i (g_{3i}(\alpha,b,\varepsilon)) \right)$$

4.4 by the derivatives above

$$\alpha_i = \sum_{j=1}^{N} w_j y_j y_i \cdot x_j \cdot x_i + \beta_i = k_i + \beta_i$$

$$\sum_{i=1}^{N} w_i(-y_i) = 0 \quad , w_i = C_i - \gamma_i$$

$L(\alpha, b, \varepsilon, w, \beta, \gamma)$

$$= \frac{1}{2}\sum_{i=1}^{N}\alpha_i^2 + \cancel{\sum_{i=1}^{N} C_i \varepsilon_i} + \sum_{i=1}^{N} w_i - \cancel{\sum_{i=1}^{N} w_i \varepsilon_i}$$
$$- \sum_{i=1}^{N} w_i \underbrace{y_i}_{=0} \left(\sum_{j=1}^{N} \alpha_j y_j x_i \cdot x_j + b\right)$$
$$+ \sum_{i=1}^{N}\beta_i(-\alpha_i) + \cancel{\sum_{i=1}^{N} \gamma_i(-\varepsilon_i)}$$

$$= \frac{1}{2}\sum_{i=1}^{N}(k_i+\beta_i)^2 - \sum_{i=1}^{N}\beta_i(k_i+\beta_i) - 0 + \sum_{i=1}^{N} w_i$$

$$= \frac{1}{2}\sum_{i=1}^{N}(k_i^2 + \cancel{2k_i\beta_i} + \beta_i)^2 - \sum_{i=1}^{N}\beta_i(k_i+\beta_i) - 0 + \sum_{i=1}^{N} w_i$$

hw5 4.4 $= \sum_{i=1}^{N} w_i - \frac{1}{2} \sum_{i=1}^{N} (\sum_{j=1}^{N} w_j y_j y_i x_j \cdot x_i)^2$

subject to $\sum_{i=1}^{N} w_i y_i = 0$

$C_i = w_i + \gamma_i \Rightarrow 0 \le w_i \le C_i$

4.5 By the derivatives in 4.3,

(a) $\bar{\alpha}_i = \sum_{j=1}^{N} \bar{w}_j y_j y_i x_j \cdot x_i + \bar{\beta}_i$

and $-\alpha_i \beta_i = 0$ for all i, $\beta_i \ge 0, \alpha_i \ge 0$

to maximize α_i, β_i must $= 0$, $\beta_i \ne 0$ only when $\alpha_i = 0$

$\bar{\alpha}_i = \max(\underbrace{0}_{\beta_i \ne 0}, \underbrace{\sum_{j=1}^{N} \bar{w}_j y_j y_i x_j \cdot x_i}_{\alpha_i \ne 0, \beta_i = 0})$

(b) (c) $1 - \varepsilon_i - y_i (\bar{w} \cdot x_j + \bar{b}) \le 0$

$\Rightarrow \varepsilon_i \ge 1 - y_i (\bar{w} \cdot x_i + b)$

and $\varepsilon_i \ge 0$, to minimize $\sum_{i=1}^{N} C_i \varepsilon_i$

$\Rightarrow \varepsilon_i = \max (1 - y_i(\bar{w} \cdot x_i + b), 0)$

(b) to minimize $\sum_{i=1}^{N} C_i \varepsilon_i$, and $\varepsilon_i = \max (1 - y_i(\bar{w} \cdot x_i + b), 0)$

$b = \underset{b \in \mathbb{R}}{\operatorname{argmin}} (\sum_{i=1}^{N} C_i \times \max(1 - y_i(\bar{w} \cdot x_i + b), 0))$

(d) $y_i(\bar{w} \cdot x_i + b) < 1$, $\varepsilon_i > 0$, by

$\gamma_i \varepsilon_i = 0$ (KKT), $\gamma_i = 0$, and $w_i + \gamma_i = C_i$

$\Rightarrow w_i = C_i$

(d) $y_i(\bar{w} \cdot x_i + b) < 1$, $\xi_i > 0$, by $\gamma_i \xi_i = 0$, $\gamma_i = 0$, and $w_i + \gamma_i = C_i$
(KKT) $\Rightarrow w_i = C_i$

$y_i(\bar{w} \cdot x_i + b) > 1$
$\Rightarrow 1 - \xi_i - y_i(\bar{w} \cdot x_i + b)$
$= -\xi_i - k$ ($\xi_i \geq 0$) (k>0)

by KKT
$w_i(1 - \xi_i - y_i(\bar{w} \cdot x_i + b)) = 0$
$\Rightarrow w_i = 0$

$y_i(\bar{w} \cdot x_i + b) = 1$
$\Rightarrow \xi_i = 0$, $0 \leq \gamma_i \leq C_i \Rightarrow 0 \leq w_i \leq C_i$ (just like original)

Math5

hw5 5.1 $L(\rho, \mu, \xi, \alpha, \beta, \gamma)$
$$= \rho + \frac{1}{\nu}\sum_{i=1}^{N} C_i \xi_i + \sum_{i=1}^{N} \alpha_i (\|x_i - \mu\|^2 - \rho - \xi_i) + \sum_{i=1}^{N} \beta_i(-\xi_i) + \gamma(-\rho)$$

5.2 domain of variable's are all convex $(R \times R^M \times R^N)$, objective function
$= \underbrace{\rho + \frac{1}{v}\sum_{i=1}^{N} C_i \xi_i}_{\text{affine}}$ is convex (linear combination)

$(\underbrace{\|x_i - \mu\|^2}_{\substack{\text{2-norm} \\ \text{(convex)}}} - \underbrace{\rho - \xi_i}_{\substack{\text{affine} \\ \text{(convex)}}})$ is convex, $-\xi_i$, $-\rho$ are convex, simply set $\rho > 0$, $\xi_i > 0$, and $\|\rho\|$ very big, we get a simple strictly feasible point, no equality constraints so $0 \in \text{int}h(0)$, slater's condition holds $\Rightarrow$ duality gap $= 0$

hw 5

5.3 $\theta(\alpha,\beta,\gamma) = \inf_{\rho,\mu,\varepsilon} \left(f(\rho,\mu,\varepsilon) + \sum_{i=1}^{N} \alpha_i g_{1i}(\rho,\mu,\varepsilon) + \sum_{i=1}^{N} \beta_i g_{2i}(\rho,\mu,\varepsilon) + \gamma g_3(\rho,\mu,\varepsilon) \right)$

5.4 $\hat{\alpha}_i = \frac{\alpha_i}{\|\alpha\|_1}$

$$\frac{\partial L}{\partial \rho} = 1 - \sum_{i=1}^{N} \alpha_i - \gamma = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i + \gamma = 1$$

$$\frac{\partial L}{\partial \varepsilon_i} = \frac{1}{v} C_i - \alpha_i - \beta_i = 0 \Rightarrow \alpha_i + \beta_i = \frac{1}{v} C_i$$

$$\frac{\partial L}{\partial \mu} = \frac{\partial}{\mu} \sum_{i=1}^{N} \alpha_i \|x_i - \mu\|^2$$

$$= \sum_{i=1}^{N} \alpha_i (2\mu - 2x_i) = 0$$

$$\rho + \frac{1}{v} \sum_{i=1}^{N} C_i \varepsilon_i$$

5-5(a) by the derivatives in 5.4

$$\sum_{i=1}^{N} \alpha_i (2\mu - 2x_i) = 0$$

$$\sum_{i=1}^{N} \alpha_i \mu = \sum_{i=1}^{N} \alpha_i x_i$$

$$\|\alpha\| \bar{\mu} = \sum_{i=1}^{N} \alpha_i x_i$$

(b) $\xi_i \geq \|x_i - \mu\|^2 - \rho$ and $\xi_i \geq 0$

to minimize $\sum_{i=1}^{N} C_i \xi_i$; $\xi_i = \max(\|x_i - \mu\|^2 - \rho, 0)$

so ρ becomes

$$\rho \in \operatorname*{argmin}_{\rho \in R} (f)$$

$$= \operatorname*{argmin}_{\rho \in R} \left(\rho + \frac{1}{V} \sum_{i=1}^{N} C_i \max(\|x_i - \mu\|^2 - \rho, 0) \right)$$

(c) since $\rho \in \operatorname{argmin} \left(\rho + \frac{1}{V} \sum_{i=1}^{N} C_i (\max \|x_i - \mu\|^2 - \rho, 0) \right)$

$$\frac{\partial h}{\partial \rho} = 1 - \frac{1}{V} \sum_{i : \|x_i - \mu\|^2 \leq \rho} C_i = 0$$

$$\bar{\rho} = \left\{ \rho \,\middle|\, \sum_{i : \|x_i - \mu\|^2 \leq \rho} C_i = V \right\}$$ this means ρ is to big, thus $\sum C_i$ become larger

obviously

$$\min \left\{ \rho \geq 0, \sum_{i : \|x_i - \mu\|^2 \leq \rho}^{N} C_i \geq V \right\} \leq \bar{\rho} \leq \left\{ \rho \geq 0 : \sum_{i : \|x_i - \mu\|^2 \leq \rho}^{N} C_i > V \right\}$$

the minmum happens when $\sum_{i : \|x_i - \mu\|^2 \leq \rho} C_i = V$, $\bar{\rho} = \rho$

5.5 (d) $\xi_i \geq 0$ and

$\xi_i \geq \|x_i - \mu\|^2 - \rho$, to minimize $\rho + \frac{1}{V}\sum_{i=1}^{N} C_i \xi_i$

$\Rightarrow \xi_i = \max(\|x_i - \mu\|^2 - \rho, 0)$

(e) if $\|x_i - \bar{\mu}\|^2 > \bar{\rho}$, which means $\xi_i > 0$

but $\beta_i(-\xi_i) = 0$ (KKT, complementary slackness)

$\Rightarrow$ so $\beta_i = 0$, and by derivatives in 5.4

we know $\bar{\alpha}_i + \bar{\beta}_i = \frac{C_i}{V}$, so $\bar{\alpha}_i = \frac{C_i}{V}$

$\|x_i - \mu\|^2 < \bar{\rho}$, which means

$\|x_i - \mu\|^2 - \rho - \xi_i \leq 0$, by complementary slackness $\bar{\alpha}_i(\|x_i - \bar{\mu}\|^2 - \bar{\rho} - \bar{\xi}_i) = 0$

$\Rightarrow \bar{\alpha}_i = 0$ $\underset{0}{\nparallel}$

$\|x_i - \bar{\mu}\|^2 = \bar{\rho}$, $\xi_i = 0$, which means

$\alpha_i + \beta_i = \frac{C_i}{V}$, $\beta_i \geq 0$, $\alpha_i \geq 0$

$\Rightarrow 0 \leq \alpha_i \leq \frac{C_i}{V}$

↳ can be any number $0 \leq \beta_i \leq \frac{C_i}{V}$