Exercise 3

1)

$$\min_{w,b} \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{n} \max\{|y_i - (w^T x_i + b)| - \varepsilon, 0\}$$

By mimicking standard SVM, can remove max and get:

$$\min_{w,b} \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{N} \gamma_i, \text{ s.t. } |y_i - (w^T x_i + b)| \le \varepsilon + \gamma_i, \gamma_i \geq 0$$

By removing absolute value sign can get:

$$\min_{w,b} \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{N} (\gamma_i, \hat{\gamma}_i)$$

$$\text{s.t. } y_i - (w^T x_i + b) \le \varepsilon + \gamma_i,$$
$$(w^T x_i + b) - y_i \le \varepsilon + \hat{\gamma}_i,$$
$$\gamma_i \geq 0, \hat{\gamma}_i \geq 0$$

let $\hat{y}_i = w^T x_i + b$ and
convert above optimal question to lagrangian dual question get:

$$\max_{\alpha,\hat{\alpha},\beta,\hat{\beta}} \min_{w,b,\gamma,\hat{\gamma}} \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{n} (\gamma_i + \hat{\gamma}_i) - \sum_{i=1}^{n} \beta_i \gamma_i - \sum_{i=1}^{n} \hat{\beta}_i \hat{\gamma}_i$$

$$+ \sum_{i=1}^{n} \alpha_i (y_i - \hat{y}_i - \gamma_i - \varepsilon) + \sum_{i=1}^{n} \hat{\alpha}_i (\hat{y}_i - y_i - \hat{\gamma}_i - \varepsilon)$$

solve $\min_{w,b,\gamma,\hat{\gamma}}$ first:

$$\frac{d}{dw} = w - \sum_{i=1}^{n} (\alpha_i - \hat{\alpha}_i) x_i = 0 \Rightarrow w = \sum_{i=1}^{n} (\alpha_i - \hat{\alpha}_i) x_i \quad ①$$

$$\frac{d}{db} = \sum_{i=1}^{n} (\alpha_i - \hat{\alpha}_i) = 0 \quad ② \quad \frac{d}{d\gamma_i} = 0 \Rightarrow C - \alpha_i - \beta_i = 0 \Rightarrow \beta_i = C - \alpha_i$$

$$\frac{d}{d\hat{\gamma}_i} = 0 \Rightarrow C - \hat{\alpha}_i - \hat{\beta}_i = 0 \Rightarrow \hat{\beta}_i = C - \hat{\alpha}_i$$

It is known that max min $L \le$ min max L,
to satisfy that max min L = min max L (complementary slackness)

$$\alpha_i(y_i - \hat{y}_i - \gamma_i - \varepsilon) = \hat{\alpha}_i(\hat{y}_i - y_i - \hat{\gamma}_i - \varepsilon) = 0 \quad ③$$

$$\beta_i\gamma_i = (C-\alpha_i)\gamma_i = \hat{\beta}_i\hat{\gamma}_i = (C-\hat{\alpha}_i)\hat{\gamma}_i = 0 \quad ④$$

put ④ into dual question get:

$$\max_{\alpha,\hat{\alpha}} \min_{w,b,\gamma,\hat{\gamma}} \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{n}(\gamma_i+\hat{\gamma}_i) - \alpha_i\sum_{i=1}^{n}\gamma_i - \hat{\alpha}_i\sum_{i=1}^{n}\hat{\gamma}_i + \sum_{i=1}^{n}(\hat{\alpha}_i-\alpha_i)\hat{y}_i$$
$$+\sum_{i=1}^{n}\left[\alpha_i(y_i-\varepsilon)+\hat{\alpha}_i(-\varepsilon-y_i)\right]$$

By ④ we know $\gamma_i = \hat{\gamma}_i = 0$ or $\alpha_i = \hat{\alpha}_i = C$

for both condition, $C\sum_{i=1}^{n}(\gamma_i+\hat{\gamma}_i) - \alpha_i\sum_{i=1}^{n}\gamma_i - \hat{\alpha}_i\sum_{i=1}^{n}\hat{\gamma}_i = 0$.

---

so we can get

$$\max_{\alpha,\hat{\alpha}} \min_{w,b} \frac{1}{2}\|w\|_2^2 + \sum_{i=1}^{n}(\hat{\alpha}_i-\alpha_i)\hat{y}_i + \sum_{i=1}^{n}\left[\alpha_i(y_i-\varepsilon)+\hat{\alpha}_i(-\varepsilon-y_i)\right]$$

put ①② into it

$$\Rightarrow \max_{\alpha,\hat{\alpha}} \min_{w,b} \frac{1}{2}\|w\|_2^2 - w^T w + \sum_{i=1}^{n}\left[\alpha_i(y_i-\varepsilon)+\hat{\alpha}_i(-\varepsilon-y_i)\right]$$
$$\text{s.t. } \sum_{i=1}^{N}(\alpha_i-\hat{\alpha}_i) = 0,\ 0\le\alpha_i\le C,\ 0\le\hat{\alpha}_i\le C$$

$$\Rightarrow \max_{\alpha,\hat{\alpha}} -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i-\hat{\alpha}_i)(\alpha_j-\hat{\alpha}_j)x_i^T x_j + \sum_{i=1}^{n}\left[\alpha_i(y_i-\varepsilon)+\hat{\alpha}_i(-\varepsilon-y_i)\right]$$
$$\text{s.t. } \sum_{i=1}^{N}(\alpha_i-\hat{\alpha}_i) = 0,\ 0\le\alpha_i\le C,\ 0\le\hat{\alpha}_i\le C$$

$$\Rightarrow \min_{\alpha,\hat{\alpha}} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i-\hat{\alpha}_i)(\alpha_j-\hat{\alpha}_j)x_i^T x_j + \sum_{i=1}^{n}\left[\alpha_i(\varepsilon-y_i)+\hat{\alpha}_i(\varepsilon+y_i)\right]$$
$$\text{s.t. } \sum_{i=1}^{N}(\alpha_i-\hat{\alpha}_i) = 0,\ 0\le\alpha_i\le C,\ 0\le\hat{\alpha}_i\le C$$

2)

$$\frac{d}{dw} = C\sum_{i=1}^{n} f(x_i, y_i, w, b), \quad f(x_i, y_i, w, b) = \begin{cases} -\text{sign}(y_i - (w^T x_i + b)) \cdot x_i, & |y_i - (w^T x_i + b)| \geq \varepsilon \\ 0, & \text{otherwise.} \end{cases}$$

$$\frac{d}{db} = C\sum_{i=1}^{n} f(x_i, y_i, w, b), \quad f(x_i, y_i, w, b) = \begin{cases} -\text{sign}(y_i - (w^T x_i + b)), & |y_i - (w^T x_i + b)| \geq \varepsilon \\ 0, & \text{otherwise.} \end{cases}$$

3)

$$\frac{d}{dz} = \frac{1}{\eta}(z - w) + z = 0$$

$$\Rightarrow \frac{z}{\eta} - \frac{w}{\eta} + z = 0$$

$$\Rightarrow \frac{(\eta+1)z}{\eta} = \frac{w}{\eta} \Rightarrow z = \frac{w}{\eta+1}$$

therefore, $P^{\eta}(w) = \dfrac{w}{\eta+1}$

4)

train error is 633.41620795
train loss is 634.18181599
test error is 794.04613278