8.1

(a) $g(x) = (x-x_a)^T(x-x_a)$

$F(x) = g(x) - \lambda f(x)$

$\frac{\partial f(x)}{\partial x} = 2x - 2x_a - \lambda w$

set $\frac{\partial f(x)}{\partial x} = 0$

$\Rightarrow \quad 2x-2x_a-\lambda w = 0 \quad ①$

$\frac{\partial f(x)}{\partial \lambda} = w^T x + b$

set $\frac{\partial f(x)}{\partial \lambda} = 0 \Rightarrow \quad w^T x + b = 0 \quad ②$

$w^T ① \Rightarrow \quad 2w^T(x-x_a) - \lambda w^T w = 0$

$\lambda = \frac{2w^T(x-x_a)}{w^T w}$

$(x-x_a)^T ① \Rightarrow \quad 2(x-x_a)^T(x-x_a) - \lambda w^T(x-x_a) = 0$

$(x-x_a)^T(x-x_a) = \frac{\frac{2w^T(x-x_a)\cdot w^T(x-x_a)}{w^T w}}{2}$

$= \frac{(w^T(x-x_a)+b-b)(w^T(x-x_a)+b-b)}{w^T w}$

$= \frac{|f(x_a)|^2}{|w|^2}$

(b) Set $x_a = 0$

$\Rightarrow \quad dis = \frac{|b|}{||w||}$

(c) $x_p = x_a - \underbrace{\frac{|f(x_a)|}{||w||}}_{distance} \cdot \underbrace{\frac{w}{||w||}}_{direction}$

$= x_a - \frac{|f(x_a)|}{||w||^2} \cdot w$

9.2

(b) $\frac{\partial L}{\partial w} = w - \sum_i^n \alpha_i y_i x_i$

set $\frac{\partial L}{\partial w} = 0$

$\Rightarrow \quad w^* = \sum_i \alpha_i y_i x_i$

$\frac{\partial L}{\partial b} = \quad - \sum_i \alpha_i y_i$

set $\frac{\partial L}{\partial b} = 0$

$\Rightarrow \quad \sum_i \alpha_i y_i = 0$

(c) $L(\alpha) = \frac{1}{2}\left(\sum_i \alpha_i y_i x_i\right)^T \left(\sum_i \alpha_i y_i x_i\right) + \sum_i \alpha_i - \sum_i \alpha_i y_i \left(\sum_j \alpha_j y_j x_j\right)^T x_i$

$= \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i y_i \alpha_j y_j x_i^T x_j$

s.t. $\sum_i \alpha_i y_i = 0$

$\alpha_i \geq 0, \ \forall i$

9.3 (a) $b = \frac{1}{y_i} - w^T x_i$

$\sum_{i \in SV} b = \sum_{i \in SV} \left(\frac{1}{y_i} - w^T x_i\right)$

$|SV| \cdot b = \sum_{i \in SV} (y_i - w^T x_i)$

$b = \frac{1}{|SV|} \sum_{i \in SV} (y_i - w^T x_i)$

(b) $b^* = \frac{1}{2} \left(1 - w^T x^+ + (-1) - w^T x^-\right)$

$= -\frac{1}{2} w^T (x^+ + x^-)$

9.4 (b) $\frac{\partial}{\partial w}L = w - \sum_i \alpha_i y_i x_i$

set $\frac{\partial}{\partial w}L = 0$

$\Rightarrow \quad w^* = \sum_i \alpha_i y_i x_i$

$\frac{\partial}{\partial b} = \sum_i \alpha_i y_i b$

set $\frac{\partial}{\partial b} = 0$

$\Rightarrow \sum_i \alpha_i y_i = 0$

$\frac{\partial}{\partial \xi_i} = C - \alpha_i - r_i$

set $\frac{\partial}{\partial \xi_i} = 0 \Rightarrow r_i = C - \alpha_i$

(c) ✓

(d) ✓

(e) discuss $\underline{\alpha_i > 0,\ \alpha_i = 0,\ r_i > 0,\ r_i = 0}$

9.5

(a) We have $\xi_i \geq 0$ ①

moreover $y_i(w^T x_i + b) \geq 1 - \xi_i$

$\Rightarrow \quad \xi_i \geq 1 - y_i(w^T x_i + b)$ ②

Combining ① and ②

$\Rightarrow \quad \xi_i \geq \max\{0, 1 - y_i(w^T x_i + b)\}$

Since the primal objective is

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C \sum_i \xi_i$$

Hence at optimal primal ξ_i will shrink as small as possible such that

$$\xi_i = \max\{0, 1 - y_i(w^T x_i + b)\}$$

(b) By (a), we know optimal ξ^* satisfies

$$\xi_i^* = \max\{0, 1 - y_i(w^T x_i + b)\}$$

Hence the original objective is equivalent to

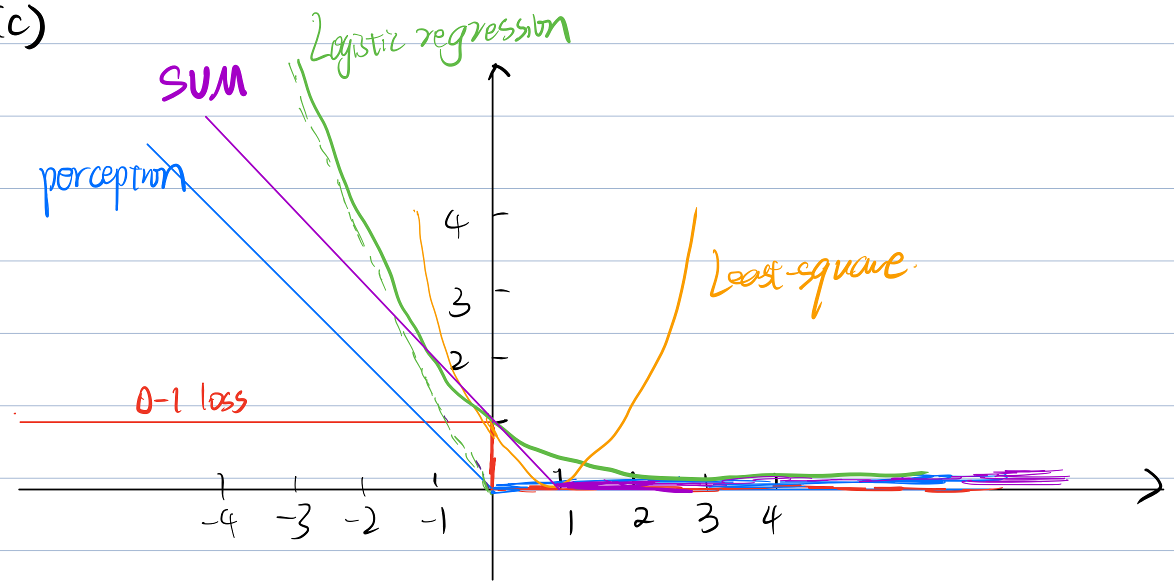
$$\min_{w,b} \frac{1}{2}\|w\|^2 + C \sum_i^n \max(0, 1 - y_i(w^T x_i + b))$$

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

(setting $\frac{1}{2C} = \lambda$)

$$= \min_{w,b} \lambda \|w\|^2 + \sum_i^n \max(0, 1 - y_i(w^T x_i + b))$$

(c)



① When $z_i < 0$ (wrong classification), the penalty linearly increases as z_i decreases. Compared with least-square, it is more robust.

② More over when $0 \le z_i \le 1$, SVM Loss function also penalize those who violate the margin which is better than perceptron.

③ when $z_i > 1$, its loss is 0 which is better than logistic regression which is slightly larger than 0

④ 0-1 loss cannot be optimized.

8.6

(a) suppose $\xi_i < 0$

and $y_i(w^T x_i + b) \geq 1 - \xi_i, \forall i$ satisfies

We have $\frac{1}{2}\|w\|^2 < \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i^2$

which also satisfies the constraint.

Hence, at the optimal solution, $\xi_i \geq 0, \forall i$ can be dropped.

(b) ✓

(c) $\frac{\partial L}{\partial w} = w - \sum_i^n \alpha_i y_i x_i$

$\Rightarrow w^* = \sum_i^n \alpha_i y_i x_i$

$\frac{\partial L}{\partial b} = -\sum_i \alpha_i y_i b$

$\Rightarrow \sum_i \alpha_i y_i = 0$

$\frac{\partial L}{\partial \xi_i} = C\xi_i - \alpha_i$

$\Rightarrow \xi_i = \frac{\alpha_i}{C}$

(d) $L(\alpha) = \sum_i^n \alpha_i - \frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j x_i^T x_j - \frac{1}{2}\sum_i \frac{\alpha_i^2}{C}$

$= \sum_i^n \alpha_i - \frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j (x_i^T x_j + \frac{1}{C}\delta_{ij})$

where $\delta_{ij} = \begin{cases} 1, & i=j \\ 0, & i\neq j \end{cases}$

(e) It will not make α_i to big as $\frac{\alpha_i^2}{C}$ regularize it.

9.8

(a) $L(w,a) = f(x) - \lambda g(x)$

$= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i (y_i f_0(x_i) + y_i w^T x_i - 1)$

(b) $\frac{\partial L}{\partial w} = w - \sum_{i=1}^{n} \alpha_i y_i x_i$

$\Rightarrow \quad w^* = \sum_i \alpha_i y_i x_i$

(c) $L(\alpha) = \frac{1}{2}\sum_i\sum_j \alpha_i y_i \alpha_j y_j x_i^T x_j$

$+ \sum_i (1 - y_i f_\theta(x))\alpha_i - \sum_i\sum_j \alpha_i y_i \alpha_j y_j x_i^T x_j$

$= \sum_i (1 - y_i f_\theta(x))\alpha_i - \frac{1}{2}\sum_i\sum_j \alpha_i y_i \alpha_j y_j x_i^T x_j$

$(\alpha_i \geq 0, \forall i)$

(d) Control the weight of α_i.

9.9

(c) $\frac{\partial L}{\partial w} = w - \sum_i (\alpha_i x_i - \hat{\alpha}_i x_i)$

let $\frac{\partial L}{\partial w} = 0$

$\Rightarrow w^* = \sum_i (\alpha_i - \hat{\alpha}_i) x_i$

$\frac{\partial L}{\partial b} = -\sum_i (\alpha_i - \hat{\alpha}_i) b$

let $\frac{\partial L}{\partial b} = 0$

$\Rightarrow \sum_i (\alpha_i - \hat{\alpha}_i) = 0$

(d) $L(\alpha,\hat{\alpha}) = \min_{w,b} L(w,b,\alpha,\hat{\alpha})$

$$= \sum_i y_i(\alpha_i-\hat{\alpha}_i) - \varepsilon\sum_i(\alpha_i+\hat{\alpha}_i)$$
$$- \frac{1}{2}\sum_i\sum_j(\alpha_i-\hat{\alpha}_i)(\alpha_j-\hat{\alpha}_j)x_i^T x_j$$

$$(\alpha_i \geq 0,\ \hat{\alpha}_i \geq 0,\ \forall i)$$
$$\sum_i \alpha_i = \sum_i \hat{\alpha}_i$$

(e) $\alpha_i = 0$, $\hat{\alpha}_i = 0$ $(\lambda = 0)$

① $\Rightarrow$ $\varepsilon - y_i + w^T x_i + b > 0$
$\varepsilon + y_i - (w^T x_i + b) > 0$

within

② ③ each boundary $g_1(x) = 0 / g_2(x) = 0$

9.10

(a) $\xi_i \geq \sum_i |y_i - f(x_i)|_\varepsilon$
$\hat{\xi}_i \geq \sum_i |y_i - f(x_i)|_\varepsilon$.
By restriction in (9.43)

When we compute the minimum of
$$\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^n(\xi_i + \hat{\xi}_i)$$
The opt will always adopt equality.

(b)

Primal $L(w, b, \xi_i, \hat{\xi}_i, \alpha_i, \hat{\alpha}_i, \lambda, \hat{\lambda}) = \frac{1}{2}\|w\|^2 + C\sum_i (\xi_i + \hat{\xi}_i)$

$- \sum_i \alpha_i(\varepsilon + \xi_i + w^T x_i + b - y_i)$

$- \sum_i \hat{\alpha}_i(\varepsilon + \hat{\xi}_i - w^T x_i - b + y_i)$

$- \sum_i \lambda_i \xi_i - \sum_i \hat{\lambda}_i \hat{\xi}_i$

$\frac{\partial L}{\partial w} = w - \sum_i x_i (\alpha_i - \hat{\alpha}_i)$

$\Rightarrow w^* = \sum_i x_i (\alpha_i - \hat{\alpha}_i)$

$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_i^n \alpha_i = \sum_j^n \hat{\alpha}_j$

$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \lambda_i \Rightarrow \lambda_i = C - \alpha_i$

$\frac{\partial L}{\partial \hat{\xi}_i} = C - \hat{\alpha}_i - \hat{\lambda}_i \Rightarrow \hat{\lambda}_i = C - \alpha_i$

$\mathcal{L}(\alpha, \hat{\alpha}) = \sum_i y_i(\alpha_i - \hat{\alpha}_i) - \varepsilon\sum_i(\alpha_i + \hat{\alpha}_i)$

$- \frac{1}{2}\sum_i\sum_j (\alpha_i - \hat{\alpha}_i)(\alpha_j - \hat{\alpha}_j) x_i^T x_j$

s.t. $0 \le \alpha_i \le C$, $0 \le \hat{\alpha}_i \le C$, $\forall i$

$\sum \alpha_i = \sum_i \hat{\alpha}_i$

(c) ① $\alpha_i = \hat{\alpha}_i = 0 \Rightarrow \underbrace{\lambda_i > 0, \hat{\lambda}_i > 0}_{\Rightarrow \xi_i = \hat{\xi}_i = 0}$

inside ε-tube