(1) (a) $x \in \mathbb{R}^d,\ y \in \{1,\dots,C\}$

$k_{struct}((x,y),(x',y')) = k(x,x') \cdot I(y=y')$

Set $M_\ell := \{i \in \{1,\dots,N\} \mid y_i = \ell\}$, for $\ell \in \{1,\dots,C\}$.

$$\sum_{i,j=1}^{N} c_i c_j k_{struct}((x_i,y_i),(x_j,y_j)) \cdot I(y_i = y_j) = \sum_{\ell=1}^{C} \underbrace{\sum_{i,j \in M_\ell} c_i c_j k(x_i,x_j)}_{\geq 0,\ \text{since } k \text{ is p.s.d.}} \geq 0$$

(b) Writing $I(y=y') = \sum_{\ell=1}^{C} I(y=\ell) \cdot I(y'=\ell)$, we see that

~~$\phi_{struct}(x,y) :=$~~

$\phi_i := \phi_{input}$

$$k_{struct}((x,y),(x',y')) = \langle \phi_i(x), \phi_i(x') \rangle \sum_{\ell=1}^{C} I(y=\ell) \cdot I(y'=\ell)$$

$$= \sum_{\ell=1}^{C} \langle I(y=\ell)\phi_i(x), I(y'=\ell)\phi_i(x') \rangle$$

→ If $\mathcal{H}$ is the RKHS for $\phi_i : \mathbb{R}^d \to \mathcal{H}$, set $\mathcal{H}_{struct} := \mathcal{H}^C = \underbrace{\mathcal{H} \times \dots \times \mathcal{H}}_{C \text{ times}}$

$\phi_{struct}(x,y) = (\phi_i(x) I(y=1), \dots, \phi_i(x) I(y=C))$, scalar product for $f = (f_1,\dots,f_C),\ g = (g_1,\dots,g_C) \in \mathcal{H}^C$ defined as

$$\langle f, g \rangle_{\mathcal{H}^C} := \sum_{\ell=1}^{C} \langle f_\ell, g_\ell \rangle.$$

(2) $\min_w \frac{1}{2}\|w\|_2^2 + C \sum_{n=1}^{N} \xi_n$ subject to $\forall n \in \{1,\dots,N\}, y \neq y_n$: $w^T \Psi_{n,y} \geq 1 - \xi_n$, $\xi_n \geq 0$,

where $\Psi_{n,y} = \phi(x,y_n) - \phi(x,y)$

(a) $$\mathcal{L} = \frac{1}{2}\|w\|_2^2 + C\sum_{n=1}^{N} \xi_n + \sum_{n=1}^{N} \sum_{y \in Y \setminus \{y_n\}} \lambda_{n,y} (w^T \Psi_{n,y} + \xi_n - 1) + \sum_{n=1}^{N} \mu_n \xi_n$$

- $\frac{\partial \mathcal{L}}{\partial w} = w + \sum_{n=1}^{N} \sum_{y \in Y \setminus \{y_n\}} \lambda_{n,y} \Psi_{n,y} \overset{!}{=} 0 \Rightarrow w = -\sum_{n=1}^{N} \sum_{y \in Y \setminus \{y_n\}} \lambda_{n,y} \Psi_{n,y}$ (I)
- $\frac{\partial \mathcal{L}}{\partial \xi_n} = C + \sum_{y \in Y \setminus \{y_n\}} \lambda_{n,y} + \mu_n \overset{!}{=} 0 \Rightarrow \mu_n = -C - \sum_{y \in Y \setminus \{y_n\}} \lambda_{n,y}$ (II)
- $\frac{\partial \mathcal{L}}{\partial \lambda_{n,y}} = w^T \Psi_{n,y} + \xi_n - 1$
- $\frac{\partial \mathcal{L}}{\partial \mu_n} = \xi_n$

Plugging (I) and (II) into the ~~de~~ equation for $\mathcal{L}$ we get: $\boxed{\text{③ (a) continued}}$

$$\frac{1}{2}\left\| -\sum_{n=1}^{N} \sum_{y \in Y \setminus \{y_n\}} \lambda_{n,y} \Psi_{n,y} \right\|_2^2 + C \sum_{n=1}^{N} \xi_n \cancel{+ \sum_{n=1}^{N} \sum_{y \in Y \setminus \{y_n\}} \lambda_{n,y} \left( \sum \right.}$$

$$+ \sum_{n=1}^{N} \sum_{y \in Y \setminus \{y_n\}} \lambda_{n,y} \left( -\sum_{n'=1}^{N} \sum_{y' \in Y \setminus \{y_{n'}\}} \lambda_{n',y'} \Psi_{n',y'}^T \Psi_{n,y} + \xi_n - 1 \right)$$

$$+ \sum_{n=1}^{N} \left( -C - \sum_{y \in Y \setminus \{y_n\}} \lambda_{n,y} \right) \xi_n$$

$$= \frac{1}{2} \sum_{n,n'=1}^{N} \sum_{y,y' \in Y \setminus \{y_n\}} \lambda_{n,y} \lambda_{n',y'} \langle \Psi_{n,y}, \Psi_{n',y'} \rangle - \sum_{n,n'=1}^{N} \sum_{y,y' \in Y \setminus \{y_n\}} \lambda_{n,y} \lambda_{n',y'} \langle \Psi_{n,y}, \Psi_{n',y'} \rangle$$

$$+ \sum_{n=1}^{N} \sum_{y \in Y \setminus \{y_n\}} \lambda_{n,y}$$

$$= \sum_{n=1}^{N} \sum_{y \neq y_n} \lambda_{n,y} - \frac{1}{2} \sum_{n,n'=1}^{N} \sum_{y,y' \neq y_n} \lambda_{n,y} \lambda_{n',y'} \langle \Psi_{n,y}, \Psi_{n',y'} \rangle$$

Since we minimize in the primal, we have to maximize now over ~~(λ)~~ $(\lambda_{n,y})_{n \in \{1,\dots,N\}, y \neq y_n}$. The constraints in the dual are

$$0 \leq \lambda_{n,y_n} \quad , n \in \{1,\dots,N\}, y_n \neq Y$$

and $0 \leq \mu_n \overset{(II)}{=} -C - \sum_{y \neq y_n} \lambda_{n,y}$, hence

$$\sum_{y \neq y_n} \lambda_{n,y} \leq -C \quad , \text{for all } n \in \{1,\dots,N\}.$$

(b) $\langle \Psi_{n,y}, \Psi_{n,y'} \rangle = \langle \phi(x,y_n) - \phi(x,y), \phi(x,y_n) - \phi(x,y') \rangle$

$$= \langle \phi(x,y_n), \phi(x,y_n) \rangle - \langle \phi(x,y), \phi(x,y_n) \rangle - \langle \phi(x,y_n), \phi(x,y') \rangle + \langle \phi(x,y), \phi(x,y') \rangle$$

$$= k((x,y_n),(x,y_n)) - k((x,y),(x,y_n)) - k((x,y_n),(x,y')) + k((x,y),(x,y'))$$

Download the programming files on ISIS and follow the instructions.

③ $L=3$

(a) $w^T \phi(x,y) = (1,1,1,1,1)^T \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ 2y_1y_2 \\ 2y_2y_3 \end{pmatrix} = y_1 - y_2 + y_3 + 2y_1y_2 + 2y_2y_3 = y_1 + (2y_1y_2 - y_2 + (2y_2y_3 + y_3))$

(b) • $\max_{y_3 \in \{-1,1\}} \{2y_2y_3 + y_3\} = \begin{cases} 1, & \text{attained for } y_3 = -1 \text{ if } y_2 = -1 \\ 3, & y_3 = 1 \text{ if } y_2 = +1 \end{cases}$

$= 2 + y_2$

$\Rightarrow \max_{y_2 \in \{\pm 1\}} 2y_1y_2 - y_2 + \max_{y_3 \in \{\pm 1\}} \{\ldots\} = \max_{y_2 \in \{\pm 1\}} 2y_1y_2 + 2$

$= \begin{cases} 4, & \text{if } y_1 = +1 = y_2 \\ 4, & \text{if } y_1 = -1 = y_2 \end{cases} = 4$

$\Rightarrow \max_{y_1 \in \{\pm 1\}} y_1 + \ldots = \max_{y_1 \in \{\pm 1\}} y_1 + 4 = 5$, attained for $y_1 = 1$

$(y_1, y_2, y_3) = (1, 1, -1)$ is the maximizer.

$\langle \ldots, \ldots \rangle = \langle \phi(x,y_\cdot) - \phi(x,y), \phi(x,y_\cdot) - \phi(x,y) \rangle$