Problem 1.a.

Solution: Let $\beta_1, ..., \beta_l \geq 0$.
The MKL kernel is defined as:

$$k(\mathbf{x}, \mathbf{x'}) = \sum_{l=1}^{L} \beta_l k_l(\mathbf{x}, \mathbf{x'}).$$

Let's write down Mercer criterion for semi definite-positiveness for MKL kernel:

$$\sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j k(\mathbf{x_i}, \mathbf{x_j}) \geq 0\,.$$

for any $x_1, ..., x_N \in \mathbb{R}^d$ and $c_1, ..., c_N \in \mathbb{R}$. Inserting the definition gives

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j \sum_{l=1}^{L} \beta_l k_l(\mathbf{x_i}, \mathbf{x_j})$$
$$= \sum_{l=1}^{L} \beta_l \left(\sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j k_l(\mathbf{x_i}, \mathbf{x_j} \right) \geq 0\,.$$

the factor between the brackets is greater or equal to zero since the k_l where defined to be positive semi-definite and because $\beta_l \geq 0$ as well, we showed that the MKL kernel $k(\mathbf{x}, \mathbf{x}'$ is positive semi-definite. □

Problem 1.b.

Solution: Let's write down the definition of MKL kernel:

$$\begin{aligned} k(\mathbf{x}, \mathbf{x'}) &= \sum_{l=1}^{L} \beta_l k_l(\mathbf{x}, \mathbf{x'}) \\ &= \sum_{l=1}^{L} \sqrt{\beta_l}\sqrt{\beta_l}\phi_l(x)^T \phi_l(x') \\ &= \sum_{l=1}^{L} (\sqrt{\beta_l}\phi_l(x)^T)(\sqrt{\beta_l}\phi_l(x')) \\ &= \sum_{l=1}^{L} (\sqrt{\beta_l}\phi_l(x))^T(\sqrt{\beta_l}\phi_l(x')) \\ &= \phi(x) \cdot \phi(x')\,. \end{aligned}$$

where the feature map of kernel k is ϕ. From the equation above we see that

$$\phi(x) = (\sqrt{\beta_1}\phi_1(x), ..., \sqrt{\beta_L}\phi_L(x))\,.$$

□

Problem 2.a.

Solution: We want to show that

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j k_{struct}((\mathbf{x_i}, y_i), (\mathbf{x_j}, y_j)) \geq 0\,,$$

for any $x_1, ..., x_N \in \mathbb{R}^d$ with their respective classes $y_1, ..., y_N \in \{1, ..., C\}$ and $c_1, ..., c_N \in \mathbb{R}$. Inserting the definition of k_{struct} gives.

$$= \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j k(x_i, x_j) \cdot 1_{[y_i=y_j]}$$

□

Problem 2.b.

Solution: Per definition of a kernel we got

$$\begin{aligned} k_{struct}((\mathbf{x}, y), (\mathbf{x}', y')) &= k(x, x') \cdot 1_{[y_i=y_j]} \\ &= \phi^T(x)\phi(x') \cdot 1_{[y_i=y_j]} \\ &= \phi^T_{struct}(\mathbf{x}, y)\phi^T_{struct}(\mathbf{x}', y')\,. \end{aligned}$$

From this we see that the feature map $\phi_{struct}(\mathbf{x}, y)$ is a C-dimensional vector (with C the number of classes) where all entries are zero except for the entry with index y which is equal to $\phi(x)$. We have here assumed that counting classes and indices starts at 1. □