Exercise Sheet 7

1 Structured Prediction for Classification

a) $\sum_{ij} c_i c_j \, k(x_i, x_j) \cdot I(y_i = y_j) \geq 0$

$= \sum c_i c_j \langle \phi(x_i), \phi(x_j) \rangle \cdot \sum_c^C I(y_i = c) \, I(c = y_j)$

$= \sum_{ij} c_i c_j \sum_\alpha \phi_\alpha(x_i) \cdot \phi_\alpha(x_j) \cdot \sum_c I(y_i = c) \, I(c = y_j)$

$= \sum_{a,c} \sum_i c_i \phi_a(x_i) I(y_i = c) \sum_j c_j \phi_a(x_j) I(c = y_j)$

$= \underbrace{\sum_{a,c} \underbrace{\Big(\sum_i c_i \phi_a(x_i) \, I(y_i = c) \Big)^2}_{\geq 0}}_{\geq 0}$

b) $k_{struct}((x,y),(x',y')) = \sum_a \phi_a(x) \phi_a(x') \cdot \sum_{c=1}^{C} I(y = c) I(y' = c)$

$= \sum_{a,c} \big(\phi_a(x) I(y = c)\big)\big(\phi_a(x') I(y' = c)\big)$

$= \left\langle \underbrace{\begin{pmatrix} \phi(x) I(y = 1) \\ \vdots \\ \phi(x) I(y = C) \end{pmatrix}}_{\phi_{struct}(x,y)}, \begin{pmatrix} \phi(x') I(y' = 1) \\ \\ \phi(x') I(y' = C) \end{pmatrix} \right\rangle$

2 Dual Formulation of Structured Output Learning

a)
$$\mathcal{L}(\cdot) = \frac{1}{2}\|w\|^2 + C\sum_n \xi_n + \sum_{n,y} \lambda_{1n,y}(1 - \xi_n - w^T\Psi_{n,y}) + \sum_n -\lambda_{2n}\xi_n$$

$$\frac{\partial \mathcal{L}(\cdot)}{\partial w} = w + \sum_{n,y}\lambda_{1n,y}(-\Psi_{n,y}) \overset{!}{=} 0 \iff w = \sum_{n,y}\lambda_{1n,y}\Psi_{n,y}$$

$$\frac{\partial \mathcal{L}(\cdot)}{\partial \xi} = C - \sum_n \lambda_{1n,y} - \lambda_{2n} \overset{!}{=} 0$$

$$\max_{\lambda_1,\lambda_2} \mathcal{L}(\cdot) = \frac{1}{2}\|\sum_{n,y}\lambda_{1n,y}\Psi_{n,y}\|^2 + C\sum_n \xi_n + \sum_{n,y}\lambda_{1n,y}$$

$$+ \sum_{n,y}\lambda_{1n,y}(-\xi_n) - \sum_{n,y}\lambda_{1n,y}\left(\sum_{n',y'}\lambda_{1n',y'}\Psi_{n',y'}\right)^T\Psi_{n,y} - \sum_n \lambda_{2n}\xi_n$$

$$= -\frac{1}{2}\left(\sum_{n,y,n',y'}\lambda_{1n,y}\lambda_{1n',y'}\Psi_{n,y}^T\Psi_{n',y'}\right) \cancel{+ C\sum_n \xi_n}$$

$$+ \underbrace{(C - \lambda_{1n,y} - \lambda_{2n})}_{=0}\sum_n \xi_n + \sum_{n,y}\lambda_{1n,y}$$

$$\lambda_{1n,y} \geq 0, \quad \lambda_{2n} \geq 0$$

$$\max_{\lambda_1} \mathcal{L}(\cdot) = \sum_{n,y}\lambda_{1n,y} - \frac{1}{2}\left(\sum_{n,y,n',y'}\lambda_{1n,y}\lambda_{1n',y'}\Psi_{n,y}^T\Psi_{n',y'}\right)$$

b)
$$\Psi_{n,y} = \phi(x, y_n) - \phi(x, y)$$
$$\Psi_{n,y}^T\Psi_{n',y'} = (\phi(x, y_n) - \phi(x, y))^T(\phi(x, y'_{n'}) - \phi(x, y'))$$

$$= \phi(x, y_n)^T\phi(x, y'_{n'}) - \phi(x, y_n)^T\phi(x, y') - \phi(x, y)^T\phi(x, y'_{n'}) + \phi(x, y)^T\phi(x, y')$$
$$= k((x, y_n)(x, y'_{n'})) - k((x, y_n), (x, y'))$$
$$- k((x, y)(x, y'_{n'})) + k((x, y), (x, y'))$$

3 Prediction of Output Sequences

a)

$$\max_y w^T \phi(x,y) = \max_y [1\ 1\ 1\ 1\ 1] \left[\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \odot \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \begin{bmatrix} 2y_1y_2 \\ 2y_2y_3 \end{bmatrix} \right]$$

$$= \max_y \mathbb{1}^T \begin{bmatrix} y_1 \\ -y_2 \\ y_3 \\ 2y_1y_2 \\ 2y_2y_3 \end{bmatrix} = \max_y y_1 - y_2 + y_3 + 2y_1y_2 + 2y_2y_3$$

$$= \max_{y_1} \left\{ y_1 + \max_{y_2} \left\{ -y_2 + 2y_1y_2 + \max_{y_3} \left\{ y_3 + 2y_2y_3 \right\} \right\} \right\}$$

b)

	$y_3=-1$	$y_3=1$
$y_2=-1$	$\boxed{1}$	-1
$\underline{y_2=1}$	-3	$\boxed{3}$

> max. for y_2

	$y_2=-1$	$y_2=1$
$y_1=-1$	4	0
$\underline{y_1=1}$	0	4

$y_1=-1$	3
$\underline{y_1=1}$	5

$\rightarrow y_1 = 1,\ y_2 = 1,\ y_3 = 1$