ML Sheet 6

1. (a) Convexity of the problem is clear since affine-linear functions and $\|\cdot\|^2$ are convex. Setting $w \in \mathbb{R}^d$ arbitrary, $\xi_1, \dots, \xi_N := 1$ and choosing $\rho < 1 + \min_{i=1,\dots,N} \langle \phi(x_i), w \rangle$ we obtain a point $(w, \xi_1, \dots, \xi_N, \rho) \in \mathbb{R}^{d+N+1} = \operatorname{int}(\mathbb{R}^{d+N+1})$ so that the inequality constraints are fulfilled with strict inequality.

(b)
$$\mathcal{L}(w, \xi, \rho, \mu, \eta) = \frac{1}{2}\|w\|^2 - \rho + \frac{1}{N\nu}\sum_{i=1}^{N} \xi_i + \sum_{i=1}^{N} \mu_i (\rho - \xi_i - \langle \phi(x_i), w \rangle) - \sum_{i=1}^{N} \eta_i \xi_i$$

(c)
- $\frac{\partial \mathcal{L}}{\partial w} = w - \sum_{i=1}^{N} \mu_i \phi(x_i) \overset{!}{=} 0$
- $\frac{\partial \mathcal{L}}{\partial \xi_i} = \frac{1}{N\nu} - \mu_i - \eta_i \overset{!}{=} 0 \quad (**)$
- $\frac{\partial \mathcal{L}}{\partial \rho} = -1 + \sum_{i=1}^{N} \mu_i \overset{!}{=} 0 \quad (*)$

Plugging into $\mathcal{L}$ yields

$$\bar{\mathcal{L}} = \frac{1}{2}\Big\|\sum_{i=1}^{N} \mu_i \phi(x_i)\Big\|^2 - \rho + \sum_{i=1}^{N} (\mu_i + \eta_i)\xi_i + \rho \cdot (=1) - \sum_{i=1}^{N} \mu_i \xi_i - \cancel{\langle w, w \rangle} - \Big\|\sum_{i=1}^{N} \mu_i \phi(x_i)\Big\|^2 - \sum_{i=1}^{N} \eta_i \xi_i$$

$$= -\frac{1}{2}\sum_{i,j=1}^{N} \mu_i \mu_j \underbrace{\langle \phi(x_i), \phi(x_j) \rangle}_{= k(x_i, x_j)}$$

The dual is maximizing $\bar{\mathcal{L}}$ over μ_i, $i = 1, \dots, N$, with constraints $0 \le \eta_i$, $0 \le \mu_i$. From $(*)$ we get $1 = \sum_{i=1}^{N} \mu_i$ and from $(**)$ it follows that $0 \le \eta_i = \frac{1}{N\nu} - \mu_i$, i.e., $\mu_i \le \frac{1}{N\nu}$.

(d) Let $A := \{ \alpha \in \mathbb{R}^N \mid \sum_{i=1}^N \alpha_i = 1,\ \forall i = 1,\dots,N: 0 \leq \alpha_i \leq \frac{1}{N\nu} \}$.
Since A is obviously compact, the continuous function $F: A \to \mathbb{R}$, $\alpha \mapsto -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j k(x_i, x_j)$

$$= -\frac{1}{2} \alpha^T K \alpha$$

attains its maximum in a point $\alpha^* \in A$.

For any $\alpha \in A$ it holds that

$$F(\alpha^*) \geq F(\alpha),$$

hence

$$-F(\alpha^*) \leq -F(\alpha).$$

This shows $\max_{\alpha \in A} F(\alpha) = \min_{\alpha \in A} -F(\alpha)$.

The condition $\sum_{i=1}^N \alpha_i = 1$ is obviously equivalent to $\mathbb{1}^T \alpha = 1$ and $\forall i = 1,\dots,N: 0 \leq \alpha_i \leq \frac{1}{N\nu}$ is equivalent to

$$\begin{pmatrix} -I \\ I \end{pmatrix} \alpha = \begin{pmatrix} -\alpha_1 \\ \vdots \\ -\alpha_N \\ \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} \preceq_K \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1/(N\nu) \\ \vdots \\ 1/(N\nu) \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbb{1}/(N\nu) \end{pmatrix}.$$

(e)