1 Hard margin SVM

Primal problem:

$$\begin{aligned}&\text{minimize } J(w) = \frac{1}{2}||w||_2^2\\ &\text{subject to } y_i(w^T x_i + w_0) \geq 1, i = \overline{1, N}\end{aligned}$$

Primal problem Lagrangian:

$$L(w_0, w, \alpha_1, \dots \alpha_N) = \frac{1}{2}||w||_2^2 - \sum_{i=1}^{N} \alpha_i [y_i(w^T x_i + w_0) - 1]$$

Karush-Kuhn-Tucker conditions:

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i$$
$$\sum_{i=1}^{N} y_i \alpha_i = 0$$
$$\alpha_i \geq 0, i = \overline{1, N}$$
$$\alpha_i [y_i (w^T x_i + w_0) - 1] = 0, i = \overline{1, N}$$
$$y_i (w^T x_i + w_0) \geq 1, i = \overline{1, N}$$

Using this we obtain the following dual problem:

$$\text{maximize } D(\alpha_1 \dots \alpha_N) = \sum_{i=1}^{N} \alpha_i - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

subject to

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$
$$\alpha_i \geq 0 i = \overline{1, N}$$

(The previous expression can be vectorized using the Hadamard matrix product and the Kronecker tensor product, done in the Python script)

2 Soft margin SVM

Primal problem:

$$
\begin{aligned}
&\text{minimize } J(w, \xi_i \dots \xi_N) = \frac{1}{2}||w||_2^2 + C\sum_{i=1}^{N}\xi_i \\
&\text{subject to } \xi_i \geq 0, y_i(w^T x_i + w_0) \geq 1 - \xi_i, i = \overline{1, N}
\end{aligned}
$$

$C > 0$ is the slackness penalty factor.

Primal problem Lagrangian:

$$
L(w, w_0, \xi_1 \dots \xi_N, \alpha_1 \dots \alpha_n, \mu_1 \dots \mu_N) = \frac{1}{2}||w||_2^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\mu_i\xi_i -
$$

$$
\sum_{i=1}^{N}\alpha_i[y_i(w^T x + w_0) - 1 + \xi_i]
$$

Karush-Kuhn-Tucker conditions:

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\mu_i \xi_i = 0, i = \overline{1, N}$$

$$\alpha_i = C - \mu_i, i = \overline{1, N}$$

$$\alpha_i [y_i(w^T x_i + w_0) - 1 + \xi_i] = 0, i = \overline{1, N}$$

$$\alpha_i \geq 0, i = \overline{1, N}$$

$$\mu_i \geq 0, i = \overline{1, N}$$

$$\xi_i \geq 0, i = \overline{1, N}$$

$$y_i(w^T x_i + w_0) \geq 1 - \xi_i, i = \overline{1, N}$$

Using this we obtain the following dual problem:

$$\text{maximize } D(\alpha_1 \dots \alpha_N) = \sum_{i=1}^{N} \alpha_i - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

subject to

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C, i = \overline{1, N}$$

The dual is essentialy the same as for the hard margin case, except for the box constraints for α_i. For kernelized version of SVM, one only needs to replace dot products with the corresponding Mercer kernel $\mathcal{K}$.