The idea of SVM is to construct boundary plane that separate the two classes of the dataset. Assume the hyperplane exists and have the form $w^T \emptyset(x) + b = 0$. We want it to be in the "middle" of two classes, which means maximizing the distance from dataset for both classes, called margin maximizing hyperplane. It separates the data to two regions, by letting $w^T \emptyset(x) + b = 1 \; or -1$, we have positive and negative hyperplanes. The margin is the distance between two planes. So it can be represent as:

$$\max \frac{2}{||w||} = \min \frac{||w||}{2}$$

We need to create a variable, ς , for the incorrect classified points that allow distance from their correct margin boundary. Turn the problem into:

$$\min \frac{1}{2} w^T w + C \sum_{i=1}^{n} \varsigma_{\mathrm{i}}$$

Subject to $y_i(w^T \emptyset(x_i) + b \geq 1 - \varsigma_{\mathrm{i}}, \; \varsigma_{\mathrm{i}} \geq 0, i = 1, \ldots n$

Where ς_{i} represent the hinge loss that can be present as $\max\,(0, 1 - \mathrm{y_i}(w^T \emptyset(x_i) + b))$ Thus we formulate the primal formulation of SVM.

However, primal formulation might not be able to use if when the data points are not linearly separable. Then we need the dual problem, transfer primal to dual:

$$\max \sum_{n} \alpha_n - \frac{1}{2} \sum_{i,j} \mathrm{y_i y_j}\; \alpha_i \alpha_j \emptyset(x_i)^T \emptyset(x_i)$$

Subject to $0 \leq \alpha_{\mathrm{n}} \leq \mathrm{C}, \mathrm{y}^{\mathrm{T}} \alpha = 0$

Here we use the kernel trick to transform data into higher dimensions by avoid explicitly computing ϕ(x). Achieve this by defining kernel function $K\left(x_i, x_j\right) = \emptyset(x_i)^T \emptyset(x_i)$

And then by letting $\mathrm{Q_{ij}} = \mathrm{y_i y_j K(x_i, x\_j)}$, we will get:

$$\min \frac{1}{2} \alpha^T Q \alpha - e^T \alpha$$

Subject to $0 \leq \alpha_{\mathrm{n}} \leq \mathrm{C}, \mathrm{y}^{\mathrm{T}} \alpha = 0$

https://github.com/ruiqi-rachel-wang/Math-253/blob/main/hw9.ipynb