Kernel Tricks.

SVM with linear boundary.

$$\max_{\alpha} \left(\sum \alpha_i - \frac{1}{2} \sum \alpha_i \alpha_j y^{(i)} y^{(j)} x^{(i)T} x^{(j)} \right)$$

$$\text{s.t.} \quad \sum \alpha_i y^{(i)} = 0, \quad \alpha_i \in \left[0, \frac{\lambda}{M}\right]$$

But linear features have their limits.
for more effective hypothesis/decision boundary
we need non-linear feature space. $x \longrightarrow \phi(x)$

e.g) $x = (1, x_1, x_2, \cdots, x_n) \in \mathbb{R}^{n+1}$

$$\begin{aligned}\phi(x) = (1, & \ x_1 x_2, x_2 x_3, \cdots, x_i x_j \cdots x_n^2 \\ & x_1 x_2 x_3, \cdots, x_i x_j x_k, \cdots, x_n^3 \\ & \vdots \\ & \underset{(p < n)}{x_1 x_2 \cdots x_p}, \quad \cdots\cdots, x_n^p)\end{aligned}$$

$\Rightarrow p + {}_nP_p$ features, very large

With large enough feature space $(\phi(x))$, it is almost impossible to solve

$$\max_{\alpha} \left(\sum \alpha_i - \frac{1}{2} \sum \alpha_i \alpha_j y^{(i)} y^{(j)} \phi(x^{(i)})^T \phi(x^{(j)}) \right)$$

$O(d^2)$ where $\phi(x) \in \mathbb{R}^d$

- Kernel function

a function $K(x^{(i)}, x^{(j)})$ that can replace $\phi(x^{(i)})^T \phi(x^{(j)})$

e.g.) $X = (x_1\ x_2\ x_3)$, $u, v \in X$

$$\phi(x) = (x_1^2, x_2^2, x_3^2, \sqrt{2}x_1x_2, \sqrt{2}x_2x_3, \sqrt{2}x_1x_3)$$

then $\phi^T(u)\phi(v) = (u_1^2v_1^2 + u_2^2v_2^2 + u_3^2v_3^2 + 2u_1u_2v_1v_2 + 2u_2u_3v_2v_3 + 2u_1u_3v_1v_3)$

$$u^Tv = (u_1v_1 + u_2v_2 + u_3v_3)$$

We define a polynomial kernel function $K(x^{(i)}, x^{(j)})$,

$K(x^{(i)}, x^{(j)}) = (x^{(i)T} x^{(j)})^p$ where p is the highest degree of the polynomial feature, then $\underbrace{\phi^T(x)\phi(v)}_{O((n+nP_2)^2)} = \underbrace{(x^T v)^2}_{O(n^2)} = K(x, v)$

the kernel function safely replaces $\phi^T\phi$ and saves us the trouble of calculating complex features.

* There are other kernel functions as well.