Lecture 23 (Non-Linear SVMs)

1 Idea

The idea is to transform the given data into a new feature space $\phi(x)$ where we then find a linear separator in this feature space.

2 Polynomial Transformation

1. $(x_1, \ldots, x_n)$ is transformed to $(\phi(x)_1, \ldots, \phi(x)_N)$ such that the tranformed polynomial is of degree d
2. The size of feature space will be $\binom{n+d}{d} = O(n^d)$

Since naive implementation is exponential in the degree, we notice the following:

$$\phi(x) = \begin{pmatrix} x_1^d \\ x_1^{d-1} x_2 \\ x_1^{d-2} x_2^2 \\ \vdots \\ \text{all possible combinations of all } x_i \\ \vdots \\ c \end{pmatrix}$$

$$\implies \phi(x)^T \phi(z) = (x^T z + c)^d$$

This can now be computed in $O(n + \log d)$ and this is called the *kernel*.

3 Kernel Function

$$K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

$$K(x, z) = \phi^T(x) \cdot \phi(z)$$

1. We can use $K(x^i, x^j)$ in the SVM dual problem and thus we can incorporate the transformation matrix easily

2. To find the separator, we can now represent it as:

$$\sum_{i=1}^{m}(\alpha_i y^i) \cdot K(x^i, x) + b$$

(b can also be computed in terms of K)
3. Effectively we only store α_i for the support vectors

3.1 Types of Kernels

1. $K(x, z) = (x^T z + c)^d$ - polynomial kernel
2. $K(x, z) = \exp\left(-\frac{||x-z||^2}{2\sigma^2}\right)$ - Gaussian kernel (transforms it to infinite dimensional space effectively)
   It is also called RBF (Radial Basis Function) kernel

4 Mercer's Theorem

Determines which functions correspond to feature transformation. Let $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a kernel function. Define:

$$K_{ij}^M = K(x^i, x^j)$$

Now, if $\exists\ \phi : \mathbb{R}^n \to \mathbb{R}^N$ such that $\phi^T(x^i) \cdot \phi(x^j) = K(x^i, x^j)$, then

1. K^M is symmetric
2. K^M is positive semi-definite

Mercer's theorem states that the converse of the above is true. Therefore, we can compute K^M and then show that there exists a feature transformation.