Foundations of Data Science & Machine Learning

Tutotial 02

March 5, 2021

Question 1. Rewrite the perceptron learning algorithm so that it works directly on any two sets of points G and B which are separated by a hyperplane (not necessarily passing through the origin). Give an upper bound on the number of updates in terms of parameters like R and δ of the G and B.

Algorithm 1 Perceptron Learning Algorithm (General Hyperplane)

Input: Two finite sets $G, B \subset \mathbb{R}^n$ which are linearly separable by a hyperplane (not necessarily passing through the origin).

Output: $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that for all $x \in G$, $\langle a, x\rangle > b$ and for all $x \in B$, $\langle a, x\rangle < b$.

```
a ← 0, b ← 0
repeat
  for all x ∈ G do
    if ⟨a, x⟩ ≤ b then
      a ← a + x
      b ← b − 1
    end if
  end for
  for all x ∈ B do
    if ⟨a, x⟩ ≥ b then
      a ← a − x
      b ← b + 1
    end if
  end for
until no updates
```

The above algorithm goes through exactly the same steps as the PLA with 1 padded to each data vector and treating b as $-a_{n+1}$. Hence this new algorithm will terminate in at most $4R^2/\delta^2$ steps where

$$
\begin{aligned}
R^2 &= \max\{\|(x_1, \ldots, x_n, 1)\|^2 : \ (x_1, \ldots, x_n) \in G \cup B\} \\
&= \max\{\|x\|^2 : \ x \in G \cup B\} + 1.
\end{aligned}
$$

and δ is the minimum distance between the convex hulls of G and B (since it is not affected by the padding).

Definition 1. A function $K : (\mathbb{R}^n \times \mathbb{R}^n) \to \mathbb{R}$ is called a *kernel* if there exists an inner product space V and a function $\phi : \mathbb{R}^n \to V$ such that

$$\forall x, y \in \mathbb{R}^n,\ K(x, y) = \langle \phi(x), \phi(y) \rangle .$$

Question 2. Show that the function $K : (\mathbb{R}^n \times \mathbb{R}^n) \to \mathbb{R}$ given by

$$K(x, y) = (1 + \langle x, y \rangle)^d$$

is a kernel for every degree $d \in \mathbb{N}$.

By considering the constant embedding $\phi(x) = \sqrt{c}, \forall x \in \mathbb{R}^n$ to $\mathbb{R}^1$, we can see that the constant function $K(x, y) = c, \forall x, y \in \mathbb{R}^n$ is a kernel. By considering the identity embedding $\phi(x) = x$, once can see that $K(x, y) = \langle x, y \rangle$ is also a kernel.

Let K_1 and K_2 be two kernels. Since they are kernels, by definition, there exists two embeddings ϕ_1 and ϕ_2 which gives K_1 and K_2. We will only prove it rigorously for the case when the co-domains of ϕ_1 and ϕ_2 are finite dimensional. Let the codomains be $\mathbb{R}^k$ and $\mathbb{R}^l$ respectively. Thus

$$\begin{aligned}
\phi_1(x) &= (f_1(x), f_2(x), \ldots, f_k(x)) \text{ and} \\
\phi_2(x) &= (g_1(x), g_2(x), \ldots, g_l(x)),
\end{aligned}$$

where f_i's and g_i's are arbitrary functions from $\mathbb{R}^n$ to $\mathbb{R}$.

We will show that $K_1 + K_2$ is a kernel. Consider the embedding $\phi : \mathbb{R}^n \to \mathbb{R}^{k+l}$ given by

$$\phi(x) = (f_1(x), f_2(x), \ldots, f_k(x), g_1(x), g_2(x), \ldots, g_l(x)).$$

It is easy to verify that $\forall x, y \in \mathbb{R}^n$

$$\begin{aligned}
\langle \phi(x), \phi(y) \rangle &= \langle \phi_1(x), \phi_1(y) \rangle + \langle \phi_2(x), \phi_2(y) \rangle \\
&= K_1(x, y) + K_2(x, y).
\end{aligned}$$

and hence $K_1 + K_2$ is a kernel.

We will show that K_1K_2 is a kernel. Consider the embedding $\phi : \mathbb{R}^n \to \mathbb{R}^{kl}$ given by

$$\begin{aligned}
\phi(x) = (&f_1(x)g_1(x), f_1(x)g_2(x), \ldots, f_1(x)g_l(x), \\
&f_2(x)g_1(x), f_2(x)g_2(x), \ldots, f_2(x)g_l(x), \\
&\ldots \\
&f_k(x)g_1(x), f_k(x)g_2(x), \ldots, f_k(x)g_l(x))
\end{aligned}$$

One can verify by expansion that

$$\begin{aligned}
K_1(x, y)K_2(x, y) =& \langle \phi_1(x), \phi_1(y) \rangle \langle \phi_2(x), \phi_2(y) \rangle \\
=&(f_1(x)f_1(y) + f_2(x)f_2(y) + \cdots + f_k(x)f_k(y))\times \\
&(g_1(x)g_1(y) + g_2(x)g_2(y) + \cdots + g_k(x)g_k(y)) \\
=& \langle \phi(x), \phi(y) \rangle .
\end{aligned}$$

When the codomains of ϕ_1 and ϕ_2 (say V and W respectively) are infinite dimensional, we can still do the $K_1 + K_2$ proof by considering ϕ as an embedding to the *direct sum* $V \oplus W$. The K_1K_2 proof is more complicated to extend and we need to consider ϕ as an embedding to the tensor product $V \otimes W$.

To complete the answer, just note that $(1 + \langle x, y \rangle)$ is a Kernel since it is the sum of two kernels and $(1 + \langle x, y \rangle)^d$ can be expressed as a repeated product of kernels.