Lecture-06: Kernel Methods

Mrugsen

Aug 21, 2018

1 RKHS Theorem Proof:

\[ \Phi(x) = \mathcal{X} \to \mathbb{R} \]
\[ x' \to K(x, x') \]
\[ i.e. \Phi_x(x') = K(x, x') \]

\( \mathbb{H} = \mathbb{H}_0 \)

Properties of the \( \langle \cdot, \cdot \rangle \):

- **Symmetric**: By the definition of PDS.
- **By PSD**, \( \langle f, f \rangle \geq 0 \)
- **Definiteness**: \( \langle f, \Phi_x \rangle^2 \leq \langle f, f \rangle \langle \Phi_x, \Phi_x \rangle \)
- **Bilinear**: \( \langle af + bg, ch + dk \rangle = ac \langle f, h \rangle + ad \langle f, k \rangle + bc \langle g, h \rangle + bd \langle g, k \rangle \)

Reproducing:

\[ f \in \mathbb{H}_0 : f = \sum_{i \in I} a_i \Phi_{x_i} \]

Verify: \( f(x) = \langle f(\cdot), K(x, \cdot) \rangle \)

LHS:

\[ f(x) = \sum_{i \in I} a_i \Phi_{x_i}(x) = \sum_{i \in I} a_i K(x_i, x) \]

RHS:

\[ \left\langle \sum_{i \in I} a_i \Phi_{x_i}, \Phi_x \rightangle = \sum_{i \in I} a_i K(x_i, x) \]
Normalization of PDS Kernels:

\[ K' : \mathcal{X} \times \mathcal{X} \to \mathbb{R} \quad \text{PDS Kernel} \]

\[ K(x, x') = \begin{cases} 
0 & \text{if } K(x, x') = 0 \text{ or } K'(x, x) = 0 \\
\frac{K'(x, x')}{\sqrt{K'(x, x)} \sqrt{K'(x', x')}} & \text{otherwise}
\end{cases} \]

Remarks:

\[ K(x, x) = 1 \quad \forall \ x \in \mathcal{X} \quad \{K(x, x) = 0\} \]

H.W.

\[ K'(x, x') = \exp\left(\frac{\langle x, x' \rangle}{\sigma^2}\right) \]

Then show that, \[ K(x, x') = \exp\left(\frac{-\|x - x'\|^2}{2\sigma^2}\right) \]

Lemma: Normalized PDS Kernels:

"Let \( K' \) be PDS kernel, then the normalized kernel \( K \) is also PDS."

Proof:

Without loss of generality, we can assume \( K'(x_i, x_i) \geq 0 \quad \forall \ i \in [m] \)

\[ \sum_{i,j=1}^{m} c_i c_j K(x_i, x_j) = \sum_{i,j} c_i c_j \frac{K'(x_i, x_j)}{\sqrt{K'(x_i, x_i)} \sqrt{K'(x_j, x_j)}} \]

\[ \Rightarrow \sum c_i c_j K'(x_i, x_j) \geq 0 \quad \text{Since } K' \text{ is PDS, Here is proved, Lets look some other forms.} \]

\[ \Rightarrow \sum c_i c_j \frac{\langle \Phi_{x_i}, \Phi_{x_j} \rangle}{\|\Phi_{x_i}\| \|\Phi_{x_j}\|} \]

\[ \Rightarrow \sum c_i c_j \left( \frac{\Phi_{x_i}}{\|\Phi_{x_i}\|}, \frac{\Phi_{x_j}}{\|\Phi_{x_j}\|} \right) \quad \text{which is,} \]

\[ \left\| \sum_{i=1}^{m} \frac{\Phi_{x_i}}{\|\Phi_{x_i}\|} \right\| ^2 \geq 0 \]

→ Advantages of working with kernel is that no explicit definition of \( \Phi \) is needed.

→ Advantages of working with explicit \( \Phi \) are:
• For primal method in various optimization problems.
• To derive an approximation based on \( \Phi \)
• Theoretical analysis where \( \Phi \) is more convenient.