AIML - CS 337

Lecture 14: Revisiting Kernel Methods

October 3rd, 2022

Lecturer: Abir De *Scribe: Adish, Pandurang, Balasubramanian, Pamar*

This lecture revisits the kernel formulation and the formulation of objective function in various form.

1 Recap : SVM formulation

Recall that

$$w^*_{svm} = \frac{\sum_i^{|D|} \alpha_i y_i x_i}{2\lambda}$$

This is linear in x, and the $dim(x) = dim(w) < \infty$

To generalise this, suppose we make it non-linear in x, but it's linear in some $\phi(x)$ which can be ∞-dimensional.

Previously the similarity mechanism involved $x_i^T x$. The new similarity mechanism uses the kernel formulation $K(x_i, x)$ for e.g $K(x_i, x) = e^{-||x_i - x||^2}$. Formally, this new "similarity measure" must have some properties, which are discussed later.

2 Hilbert Space and Kernel

Definition 2.1. Inner Product Space An inner product space(over reals) is a vector space V and an inner product, that is a map

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$$

that satisfies the following three properties for all vectors $x, y, z \in V$ and all scalars $a, b \in \mathbb{R}$

1. Symmetry
$$\langle x, y \rangle = \langle y, x \rangle$$

2. Linearity in the first argument
$$\langle ax + by, z \rangle = a\langle x, z \rangle + b\langle y, z \rangle$$

3. Positive definitive
$$\langle x, x \rangle \geq 0 \text{ and } \langle x, x \rangle = 0 \iff x = 0$$

Norm induced by the inner product: $||x|| = \sqrt{\langle x, x \rangle}$

Definition 2.2. Hilbert Space

A *Hilbert Space* is an inner product space that is complete and seperable with respect to the norm defined by the inner product.

Examples of Hilbert spaces include :

1. $\mathbb{R}^n$ is an Hilbert space for the Euclidean norm. The dot-product is defined as with $\langle a, b\rangle = a^T b$, the vector dot product of a and b.

2. The space l_2 of square summable sequences, with inner product $\langle x, y\rangle = \sum_{i=0}^{\infty} x_i y_i$

Definition 2.3. Kernel

Let $\mathcal{X}$ be a non-empty set. A function $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a kernel if there exists a Hilbert space $\mathcal{H}$ and a feature map $\phi : \mathcal{X} \to \mathcal{H}$ such that $\forall x, x' \in \mathcal{X}$,

$$K(x, x') = \langle \phi(x), \phi(x')\rangle_{\mathcal{H}}$$

If we are given a function of two arguments, $K(x, x')$, the following can be used to determine if it is a valid kernel.

1. Find a feature map. But this may not be obvious sometimes, and the feature map may not be unique.

2. Can use a direct property of the function which is positive definiteness. The following lemma gives a sufficient and necessary condition.

Lemma 2.4. *Let $\mathcal{H}$ be a Hilbert space, $\mathcal{X}$ a non-empty set and $\phi : \mathcal{X} \to \mathcal{H}$. A symmetric function $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ implements an inner product in $\mathcal{H}$ if and only if it is positive semidefinite; namely $\forall (x_1, \dots, x_n) \in \mathcal{X}^n$, the Gram matrix $G_{i,j} = K(x_i, x_j)$, is a positive semidefinite matrix.*

Proof. $\implies$ (If K implements an inner product then the Gram matrix is positive semidefinite)

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j K(x_i, x_j) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \langle a_i\phi(x_i), a_j\phi(x_j)\rangle_{\mathcal{H}} \\ &= \|\sum_{i=1}^{n} a_i\phi(x_i)\|_{\mathcal{H}}^2 \geq 0\end{aligned}$$

$\impliedby$ For this direction, define the space of functions over $\mathcal{X}$ as $\mathbb{R} = \{f : \mathcal{X} \to \mathbb{R}\}$ For each $x \in \mathcal{X}$ let $\phi(x)$ be the function $x \mapsto K(\cdot, x)$. Define a vector space by taking all linear combinations of elements of the form $K(\cdot, x)$. Define an inner product on this vector space to be

$$\langle \sum_i \alpha_i K(\cdot, x_i), \sum_j \beta_j K(\cdot, x'_j)\rangle = \sum_{i,j} \alpha_i \beta_j K(x_i, x'_j)$$

This is a valid inner product since it is symmetric (because K is symmetric), it is linear, and it is positive definite. Clearly,

$$\langle \phi(x), \phi(x')\rangle = \langle K(\cdot, x), K(\cdot, x')\rangle = K(x, x').$$

□

3 Generalised objective function

Consider

$$\underset{w}{min} \;\; l(\{w^T\phi(x_i)\}_{i\in D}, \{y_i\}_{i\in D}) + \lambda R(||w||) \tag{1}$$

where $l : \mathbb{R}^{|D|} \to \mathbb{R}$ is an arbitrary function and $R : \mathbb{R}_+ \to \mathbb{R}$ is a monotonically non-decreasing function.

Theorem 3.1. ***Representer Theorem***

Assume that ϕ is a mapping from $\mathcal{X}$ to a Hilbert space. Then, there exists a vector $\alpha \in \mathbb{R}^{|D|}$ such that $w = \sum_{i=1}^{|D|} \alpha_i \phi(x_i)$ is an optimal solution of equation 1

Proof. Let w^* be an optimal solution of Equation 1. Because w^* is an element of a Hilbert space, we can rewritew^* as

$$w* = \sum_{i=1}^{|D|} \alpha_i \phi(x_i) + u$$

where $\langle u, \phi(x_i)\rangle = 0$ for all i. Set $w = w^* - u$. Clearly, $||w^*||^2 = ||w||^2 + ||u||^2$, thus $||w|| < ||w^*||$. Since R is non-decreasing we obtain that $R(||w||) < R(||w^*||)$. Additionally, for all i we have that

$$\langle w, \phi(x_i)\rangle = \langle w^* - u, \phi(x_i)\rangle = \langle w^*, \phi(x_i)\rangle,$$

hence

$$l(\{w^T\phi(x_i)\}_{i\in D}, \{y_i\}_{i\in D}) = l(\{w^{*T}\phi(x_i)\}_{i\in D}, \{y_i\}_{i\in D})$$

We have shown that the objective of Equation 1 at w cannot be larger than the objective at w^* and therefore w is also an optimal solution. Since $w = \sum_{i=1}^{|D|} \alpha_i \phi(x_i)$ we conclude our proof. $\square$

Form of f is

$$\begin{aligned} f(x) &= w^{*T}\phi(x_i) \\ &= \sum_{i=1}^{|D|} \alpha_i \phi^T(x_i)\phi(x) \end{aligned}$$

Here $\phi^T(x_i)\phi(x)$ is like a similarity measure If $\phi(\cdot)$ is ∞-dimensional, we can write it as

$$f(x) = \sum_{i=1}^{|D|} \alpha_i \sum_{j=0}^{\infty} \phi(x_i)[j]\phi(x)[j]$$

Hence, if $\phi(\cdot)$ is ∞-dimensional, it is not feasible to code w as it has the same dimension as ϕ. So, we try to represent the objective function in functional form or through the kernel formulation.

3.1 Objective in terms of Kernel

Writing $w = \sum_{j=1}^{|D|} \alpha_j \phi(x_j)$, we have that for all i

$$\langle w, \phi(x_i)\rangle = \langle \sum_{j=1}^{|D|} \alpha_j \phi(x_j), \phi(x_i\rangle = \sum_{j=1}^{|D|} \alpha_j \langle \phi(x_j), \phi(x_i)\rangle.$$

Similarly,

$$||w||^2 = \langle \sum_{j=1}^{|D|} \alpha_j \phi(x_j), \sum_{j=1}^{|D|} \alpha_j \phi(x_j)\rangle = \sum_{i,j=1}^{|D|} \alpha_i \alpha_j \langle \phi(x_i), \phi(x_j)\rangle.$$

Let $K(x, x') = \langle \phi(x), \phi(x')\rangle$ be a function that implements the kernel function with respect to the feature space. Hence, instead of solving Equation 1, we can solve the equivalent problem

$$\underset{\alpha \in \mathbb{R}^{|D|}}{min}\, l(\{\sum_{j=1}^{|D|} \alpha_j K(x_j, x_i)\}_{i \in D}, \{y_i\}_{i \in D}) + \lambda R(\sqrt{\sum_{i,j=1}^{|D|} \alpha_i \alpha_j K(x_j, x_i)}) \tag{2}$$

3.2 Objective in terms of functional form

$$f(x) = \sum_i \alpha_i K(x_i, x)$$

Function f forms a vector space

$$f_1, f_2 \in V \implies af_1 + bf_2 \in V$$
$$0 \in V, \text{ by putting } \alpha_i = 0 \quad \forall i$$

Now define an inner product

$$\langle f, g\rangle_H = \sum \alpha_i^f \alpha_j^g K(x_i, x_j) \tag{3}$$

From the properties of inner product space, for 3 to be true, $\sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) \geq 0$ and $K(x_i, x_j) = K(x_j, x_i) \quad \forall i, j$

These are in accordance with the earlier lemma which we proved.

$$\begin{aligned}
||w||^2 &= \langle w, w\rangle \\
&= \langle \sum_{i=1}^{|D|} \alpha_i \phi(x_i), \sum_{i=1}^{|D|} \alpha_j \phi(x_j)\rangle = \sum_{i,j=1}^{|D|} \alpha_i \alpha_j \langle \phi(x_i) \phi(x_j)\rangle \\
&= \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) = \langle f, f\rangle \\
&= ||f||^2
\end{aligned}$$

Hence, the objective function becomes

$$\underset{w}{min}\ \ l(\{f(x_i)\}_{i \in D}, \{y_i\}_{i \in D}) + \lambda R(||f||)$$

4 Reproducing kernel Hilbert spaces

For a Hilbert space $\mathcal{H}$ of real-valued functions on $\mathcal{X}$, and for any point $x \in \mathcal{X}$, the evaluation functional at x is defined as the map $L_x : \mathcal{H} \mapsto \mathbb{R}$ such that for all functions $f \in \mathcal{H}$,

$$L_x(f) = f(x). \tag{4}$$

In this setting, $\mathcal{H}$ is called a reproducing kernel Hilbert space if for all $x \in \mathcal{X}$, L_x is bounded, i.e. there is some finite constant M such that

$$|L_x(f)| = |f(x)| \leq M\|f\|_{\mathcal{H}}. \tag{5}$$

(Equivalently, for all $x \in \mathcal{X}$, L_x is continuous at any $f \in \mathcal{H}$.)

5 Example problem

5.1 Problem Statement

Consider the functions $h : \mathbb{N} \to [1 \dots m]$ and $\mathcal{E} : \mathbb{N} \to \pm 1$

$$a^{h,\mathcal{E}}(x)[i] = \sum_{j \text{ s.t } h(i)=j} \mathcal{E}(j) x_j$$

Then prove that

$$\mathop{\mathbb{E}}_{h,\mathcal{E} \sim \ \mathcal{U}(.)}[\langle a^{h,\mathcal{E}}(x), a^{h,\mathcal{E}}(x')\rangle] = \langle x, x' \rangle$$

5.2 Solution

$$\mathop{\mathbb{E}}_{h,\mathcal{E} \sim \ \mathcal{U}(.)}[\langle a^{h,\mathcal{E}}(x), a^{h,\mathcal{E}}(x')\rangle] = \mathop{\mathbb{E}}_{h,\mathcal{E} \sim \ \mathcal{U}(.)}[\sum_{j;h(i)=j} \sum_{j';h(i)=j'} \mathcal{E}(j)\mathcal{E}(j') x_j x'_{j'}]$$

Note that since h and $\mathcal{E}$ are sampled from uniform distributions, for $j \neq j'$
$\mathbb{E}[\mathcal{E}(j)\mathcal{E}(j')] = 0$ and for $j = j'$, $\mathbb{E}[\mathcal{E}(j)\mathcal{E}(j)] = 1$

Therefore the expectation simplifies to

$$\mathop{\mathbb{E}}_{h,\mathcal{E} \sim \ \mathcal{U}(.)}[\sum_{j=j'} (1) * x_j x'_{j'} + 0] = \mathbb{E}[\sum_j x(j) x'(j)] = \mathbb{E}[\langle x, x' \rangle] = \langle x, x' \rangle$$