Workshop 5

COMP90051 Machine Learning

Semester 2, 2020

Learning Outcomes

At the end of this workshop you should be able to:

1. be able to prove the validity of kernels, e.g. by applying Mercer’s theorem
2. be able to fit SVMs in scikit-learn using a grid search for the hyperparameters
3. develop intuition about how SVM decision surfaces are affected by hyperparameters

Worksheet 5a

W5a: Kernel identities

Let K_1 and K_2 be valid kernels on a vector space $\mathcal{X}$, $c > 0$ be a constant and $f()$ be a real-valued function on $\mathcal{X}$.

Prove that the following new kernels are also valid:

- $K(\mathbf{u}, \mathbf{v}) = cK_1(\mathbf{u}, \mathbf{v})$
- $K(\mathbf{u}, \mathbf{v}) = K_1(\mathbf{u}, \mathbf{v}) + K_2(\mathbf{u}, \mathbf{v})$
- $K(\mathbf{u}, \mathbf{v}) = f(\mathbf{u})K_1(\mathbf{u}, \mathbf{v})f(\mathbf{v})$

Hint: you may use Mercer's theorem.

Mercer's theorem

A symmetric function $K: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a valid kernel on $\mathcal{X}$ if the Gram matrix

$$\mathbf{K} = \begin{bmatrix} K(x_1, x_1) & \cdots & K(x_1, x_n) \\ \vdots & \ddots & \vdots \\ K(x_n, x_1) & \cdots & K(x_n, x_n) \end{bmatrix}$$

is positive semidefinite for all finite sequences $x_1, x_2, \ldots, x_n \in \mathcal{X}$.

W5a: Kernel identities

Prove that $K(\mathbf{u},\mathbf{v}) = cK_1(\mathbf{u},\mathbf{v})$ is a valid kernel

Let ϕ_1 be the feature map associated with K_1.

Observe that:

$$K(\mathbf{u},\mathbf{v}) = cK_1(\mathbf{u},\mathbf{v}) = \left\langle \sqrt{c}\phi_1(\mathbf{u}), \sqrt{c}\phi_1(\mathbf{v}) \right\rangle$$

Since K can be written as an inner product with the feature map $\phi(\mathbf{u}) = \sqrt{c}\phi_1(\mathbf{u})$, it is a valid kernel.

W5a: Kernel identities

Prove that $K(\mathbf{u}, \mathbf{v}) = K_1(\mathbf{u}, \mathbf{v}) + K_2(\mathbf{u}, \mathbf{v})$ is a valid kernel

We'll give a proof by Mercer's theorem.

Given a set of vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n \in \mathcal{X}$ let $\mathbf{K}_1$ be the gram matrix for kernel K_1 (similarly for $\mathbf{K}_2$ and K_2)

The Gram matrix associated with K is $\mathbf{K} = \mathbf{K}_1 + \mathbf{K}_2$.

Now $\mathbf{K}$ is positive semi-definite since for any $\mathbf{x} \in \mathbb{R}^n$

$$\mathbf{x}^\top \mathbf{K} \mathbf{x} = \mathbf{x}^\top \mathbf{K}_1 \mathbf{x} + \mathbf{x}^\top \mathbf{K}_2 \mathbf{x} \geq 0$$

(since $\mathbf{K}_1$ and $\mathbf{K}_2$ are both p.s.d.)

W5a: Kernel identities

Prove that $K(\mathbf{u}, \mathbf{v}) = f(\mathbf{u}) K_1(\mathbf{u}, \mathbf{v}) f(\mathbf{v})$ is a valid kernel

Let ϕ_1 be the feature map associated with K_1.

Observe that:

$$K(\mathbf{u}, \mathbf{v}) = f(\mathbf{u}) K_1(\mathbf{u}, \mathbf{v}) f(\mathbf{v}) = \langle f(\mathbf{u}) \phi_1(\mathbf{u}), f(\mathbf{v}) \phi_1(\mathbf{v}) \rangle$$

Since K can be written as an inner product with the feature map $\phi(\mathbf{u}) = f(\mathbf{u}) \phi_1(\mathbf{u})$, it is a valid kernel.

Worksheet 5b