Workshop 7

COMP90051 Machine Learning

Semester 2, 2018

Learning Outcomes

At the end of this workshop you should:

1. be able to compare SVMs to other classification algorithms covered in this course — Discussion
2. have experience in applying Mercer's theorem (to prove the validity of a kernel) — Pen and paper
3. be able to fit SVMs in scikit-learn using a grid search for the hyperparameters — Worksheet 7

Class Discussion

Discuss the advantages and disadvantages of SVMs when compared to other classification algorithms.

Class Discussion

Advantages of SVMs

- Flexibility (non-linearity) through the kernel trick
- Reduces to a convex optimisation problem
- Robustness through regularisation/max margin

Disadvantages of SVMs

- Training scales poorly with data set size
- Uncalibrated class membership probabilities
- Interpretability
- Effectiveness depends on choice of kernel/parameters
- Not directly applicable to multi-class tasks

Valid kernels

Let K_1 and K_2 be valid kernels on a vector space $\mathcal{X}$, $c > 0$ be a constant and $f: \mathcal{X} \to \mathbb{R}$.

Prove that the following new kernels are also valid:

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

Hint: you may use Mercer’s theorem on the following slide.

Mercer's theorem

Consider a symmetric function $K(\cdot,\cdot)$ defined on a vector space $\mathcal{X}$.

K is a valid kernel 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 $x_1, x_2, \ldots, x_n \in \mathcal{X}$