Notes on hash kernels

Bart Frenk

April 20, 2018

1 Meta

1. Weinberger, Dasgupta, Langford, Smola, Attenberg. Feature hashing for large scale multitask learning. Proceedings of the 26th International Conference on Machine Learning, Montreal, Canada, 2009. (Mostly Yahoo!)
2. Shi, Petterson, Langford, Smola, Strehl, Dror, Vishnwanathan. Hash kernels. 2009.

2 Feature hashing for large scale multitask learning

Use a random function to map a higher dimensional space (of features) into a lower dimensional space. The expectation of the inner product of the range equals the inner product on the domain.

The article is sloppy with probability distributions. How to define one on the function space $\{1, \ldots, m\} \to \mathbb{R}$?.

2.1 Introduction

2.2 Hash functions

Let $\mathcal{X}$ be a high dimensional feature space, and $\mathcal{Y}$ a lower dimensional target space. Both are equipped with inner products. The aim is to define a random variable Φ with values in the function space $\mathcal{X} \to \mathcal{Y}$, such that

$$x_1 \cdot x_2 = \mathbb{E}(\Phi(x_1) \cdot \Phi(x_2)).$$

It is important that $\Phi(x_1) \cdot \Phi(x_2)$ has low variance.

3 Hash kernels

4 Experiments