Data mining - Homework 1: Random feature maps

February 10, 2019

1 Problem 1: Comparing different dimensionality reduction mechanisms (30 points)

The goal of this task is to compare different methods for dimensionality reduction. Consider the following methods: standard baseline using unstructured Gaussian matrices (IID), circulant method using Gaussian circulant matrices (CIRC), the one using Gaussian orthogonal matrices (GORT), the one applying random Hadamard matrices with three HD blocks (HD) and the one using Kac's random walk matrices built of $\lceil d \log d \rceil$ Givens random rotations (KAC). Construct two 16-dimensional vectors $\mathbf{x}, \mathbf{y}$ with $|\mathbf{x}^\top \mathbf{y}| > 1.0$ (alternatively you can sample two vectors from the dataset boston-full.npz uploaded to Courseworks and apply padding mechanism to make them 16-dimensional). Propose an algorithm to compute empirical mean-squared error (MSE) of the estimator of $\mathbf{x}^\top \mathbf{y}$ based on the above methods and present your results by plotting the empirical MSE as a function of the number random features m used. Take $m = 1, 2, 4, 6, 8, 10, 12, 14$. You can assume that HD and KAC mechanisms use sampling without repetitions. What conclusions regarding the accuracy of different estimators can you derive ?

2 Problem 1: Orthogonal estimators for nonisotropic Gaussian kernels (20 points)

Consider nonisotropic Gaussian kernel $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined as $k(\mathbf{x}, \mathbf{y}) = e^{\frac{-\tau^\top \mathbf{Q} \tau}{2}}$, where $\tau = \mathbf{x} - \mathbf{y}$ and $\mathbf{Q}$ is some positive definite matrix. Propose a method of estimating this kernel using random feature maps and based on Gaussian orthogonal matrices that can be more accurate than unstructured baseline. Justify your answer. Give time complexity of the computation of the random matrix used in that mechanism to construct your random feature maps.