Homework 2 Written Assignment

10-605/10-805: Machine Learning with Large Datasets

Due Tuesday, September 29th at 1:30:00 PM Eastern Time

Submit your solutions via Gradescope, **with your solution to each subproblem on a separate page**, i.e., following the template below. Note that Homework 2 consists of two parts: this written assignment, and a programming assignment. The written part is worth **30%** of your total HW2 grade (programming part makes up the remaining 70%).

1 Nyström Method (30 points)

Nyström method. Define the following block representation of a kernel matrix:

$$\mathbf{K} = \begin{bmatrix} \mathbf{W} & \mathbf{K}_{21}^\top \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{bmatrix} \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} \mathbf{W} \\ \mathbf{K}_{21} \end{bmatrix}.$$

The Nyström method uses $\mathbf{W} \in \mathbb{R}^{l \times l}$, $\mathbf{C} \in \mathbb{R}^{m \times l}$ and $\mathbf{K} \in \mathbb{R}^{m \times m}$ to generate the approximation $\widetilde{\mathbf{K}} = \mathbf{C}\mathbf{W}^\dagger\mathbf{C}^\top \approx \mathbf{K}$.

(a) *[5 points]* Show that $\mathbf{W}$ is symmetric positive semi-definite (SPSD) and that $\left\|\mathbf{K} - \widetilde{\mathbf{K}}\right\|_F = \left\|\mathbf{K}_{22} - \mathbf{K}_{21}\mathbf{W}^\dagger\mathbf{K}_{21}^\top\right\|_F$, where $\|.\|_F$ is the Frobenius norm.

(b) *[10 points]* Let $\mathbf{K} = \mathbf{X}^\top \mathbf{X}$ for some $\mathbf{X} \in \mathbb{R}^{N \times m}$, and let $\mathbf{X}' \in \mathbb{R}^{N \times l}$ be the first l columns of $\mathbf{X}$. Show that $\widetilde{\mathbf{K}} = \mathbf{X}^\top \mathbf{P}_{U_{X'}} \mathbf{X}$, where $\mathbf{P}_{U_{X'}}$ is the orthogonal projection onto the span of the left singular vectors of $\mathbf{X}'$.

(c) *[5 points]* Is $\widetilde{\mathbf{K}}$ symmetric positive semi-definite (SPSD)?

(d) *[5 points]* If $\text{rank}(\mathbf{K}) = \text{rank}(\mathbf{W}) = r \ll m$, show that $\widetilde{\mathbf{K}} = \mathbf{K}$. Note: this statement holds whenever $\text{rank}(\mathbf{K}) = \text{rank}(\mathbf{W})$, but is of interest mainly in the low-rank setting.

(e) *[5 points]* If $m = 20\text{M}$ and $\mathbf{K}$ is a dense matrix, how much space is required to store $\mathbf{K}$ if each entry is stored as a double? How much space is required by the Nyström method if $l = 10\text{K}$?