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.

Solution: For the first part of question, note that $\mathbf{W}$ is SPSD if $\mathbf{x}^\top\mathbf{W}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^l$. This condition is equivalent to $\mathbf{y}^\top\mathbf{K}\mathbf{y} \geq 0$ for all $\mathbf{y} \in \mathbb{R}^m$ where $y_i = 0$ for $l + 1 \leq i \leq m$. Since $\mathbf{K}$ is SPSD by assumption, this latter condition holds. For the second part, we write $\widetilde{\mathbf{K}}$ in block form as

$$\widetilde{\mathbf{K}} = \begin{bmatrix} \mathbf{W} \\ \mathbf{K}_{21} \end{bmatrix} \mathbf{W}^\dagger \begin{bmatrix} \mathbf{W} & \mathbf{K}_{21}^\top \end{bmatrix} = \begin{bmatrix} \mathbf{W} & \mathbf{K}_{21}^\top \\ \mathbf{K}_{21} & \mathbf{K}_{21}\mathbf{W}^\dagger\mathbf{K}_{21}^\top \end{bmatrix}.$$

Comparison with the block form of $\mathbf{K}$ then immediately yields the desired result.

(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 $\mathbf{X}'$.

Solution: Observe that $\mathbf{C} = \mathbf{X}^\top \mathbf{X}'$ and $\mathbf{W} = \mathbf{X}'^\top \mathbf{X}'$. Thus,

$$\widetilde{\mathbf{K}} = \mathbf{C}\mathbf{W}^\dagger \mathbf{C}^\top = \mathbf{X}^\top \mathbf{X}' \left(\mathbf{X}'^\top \mathbf{X}'\right)^\dagger \mathbf{X}'^\top \mathbf{X} = \mathbf{X}^\top \mathbf{U}_{X'} \mathbf{U}_{X'}^\top \mathbf{X} = \mathbf{X}^\top \mathbf{P}_{U_{X'}} \mathbf{X}.$$

(c) *[5 points]* Is $\widetilde{\mathbf{K}}$ SPSD?

Solution: Yes. Using the expression for $\widetilde{\mathbf{K}}$ in (b) and the idempotency of orthogonal projection matrices, we can write $\widetilde{\mathbf{K}} = \mathbf{X}^\top \mathbf{P}_{U_{X'}} \mathbf{X} = \mathbf{A}^\top \mathbf{A}$, where $\mathbf{A} = \mathbf{P}_{U_{X'}} \mathbf{X}$.

(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.

Solution: Since $\mathbf{K} = \mathbf{X}^\top \mathbf{X}$, $\text{rank}(\mathbf{K}) = \text{rank}(\mathbf{X}) = r$. Similarly, $\mathbf{W} = \mathbf{X}'^\top \mathbf{X}'$ implies $\text{rank}(\mathbf{W}) = \text{rank}(\mathbf{X}') = r$. The columns of $\mathbf{X}'$ are columns of $\mathbf{X}$ and they thus span the columns of $\mathbf{X}$. Hence, $\mathbf{U}_{X'}$ is an orthonormal basis for $\mathbf{X}$, i.e., $\mathbf{I}_N - \mathbf{P}_{U_{X'}} \in \text{Null}(\mathbf{X})$, and by part (b) of this exercise we have $\mathbf{K} - \widetilde{\mathbf{K}} = \mathbf{X}^\top (\mathbf{I}_N - \mathbf{P}_{U_{X'}})\mathbf{X} = \mathbf{0}$.

(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 $l = 10\text{K}$?

Solution: Storage of $\mathbf{K}$ requires roughly 3200 TB, i.e.,

$$(20 \times 10^6)^2 \text{ entries} \times 8 \text{ bytes/entry} \times \frac{1\text{TB}}{10^{12} \text{ bytes}} = 3200 \text{ TB}.$$

Storage of $\mathbf{C}$ requires roughly 160 GB, i.e.,

$$(20 \times 10^6 \times 10^4) \text{ entries} \times 8 \text{ bytes/entry} \times \frac{1\text{GB}}{10^9 \text{ bytes}} = 1600 \text{ GB}.$$

Note that the computed numbers do not account for the symmetry of $\mathbf{K}$ (doing so would change the storage requirements by less than a factor of two)