10-605: Machine Learning with Large Datasets

Fall 2020

Recitation 4

25th September, 2020

Quick Review - SPSD matrices

A symmetric matrix $K \in \mathbb{R}^{n \times n}$ is said to be positive semidefinite iff all its eigenvalues are non-negative.

Furthermore, for any symmetric matrix K, the following statements are equivalent:

1. $z^T K z \geq 0$ for every vector z
2. There exists a matrix A such that $A^T A = K$
3. All the eigenvalues are non-negative

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.

Note: You can assume that W is full-rank.

(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 $\mathrm{rank}(\mathbf{K}) = \mathrm{rank}(\mathbf{W}) = r \ll m$, show that $\widetilde{\mathbf{K}} = \mathbf{K}$. Note: this statement holds whenever $\mathrm{rank}(\mathbf{K}) = \mathrm{rank}(\mathbf{W})$, 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 $l = 10\text{K}$?

Additional References

Linear Algebra Review:

1. 3Blue 1Brown - https://www.youtube.com/watch?v=fNk_zzaMoSs&list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab

Projection Matrices:

1. Piazza discussion post - https://piazza.com/class/kcxot365ufc53w?cid=161
2. https://ocw.mit.edu/courses/mathematics/18-06sc-linear-algebra-fall-2011/least-squares-determinants-and-eigenvalues/projections-onto-subspaces/MIT18_06SCF11_Ses2.2sum.pdf
3. https://jekyll.math.byuh.edu/courses/m343/handouts/svdbasis.pdf