COMP4680/COMP8650: Advanced Topics in Machine Learning

Assignment #1: Background

Due: 11:59pm on Friday 5 August, 2022.
Submit solutions (with working) as a single PDF file or Jupyter notebook via Wattle.
Don't forget to include your name and student ID at the top of the submitted solutions.

1. **Sets (10 marks).** Let $A = [-1, 1)$, let $B = [0, 3]$ and let $C = [-1, 0]$. Find
   (a) $A \cap B$
   (b) $A \cup B$
   (c) $A \cap C$
   (d) $A \cup C$
   (e) $B \cap C$
   (f) $B \cup C$
   (g) $A \times B$
   (h) $\sup (A \setminus B)$
   (i) $\inf (A \cap \mathbb{R})$
   (j) $\sup (\mathbb{R} \setminus B)$
2. **Diagonal matrices (10 marks).** Let $D \in \mathbb{R}^{n \times n}$ be an n-by-n diagonal matrix.
   (a) Show that the cost of computing $y = Dx$ for any vector $x \in \mathbb{R}^n$ is $O(n)$.
   (b) Let A be an arbitrary n-by-n matrix. Does $DA = AD$? Prove or give a counterexample.
   (c) Using the formula $\det A = \sum_\sigma \operatorname{sgn}(\sigma) \prod_{i=1}^n A_{i,\sigma(i)}$ where the summation is over all permutations σ of $(1, 2, \ldots, n)$, show that the determinant of D equals the product of the diagonal entries.
   (d) Under what conditions will D be positive definite?
3. **Triangular matrices (10 marks).** Show that for upper or lower triangular matrices the eigenvalues are equal to the diagonal elements. Use this fact to give a condition for when a triangular matrix is invertible.
4. **Determinant of a transpose (10 marks).** Prove that $\det A^T = \det A$.
5. **Cauchy-Schwarz inequality (10 marks).** Given any $x \in \mathbb{R}^n$ satisfying $\sum_{i=1}^n x_i = 1$, use the Cauchy-Schwarz inequality to prove that $\|x\| \geq \frac{1}{\sqrt{n}}$. Does this lower bound change if we further restrict x to be non-negative, i.e., $x_i \geq 0$ for $i = 1, \ldots, n$ and $\sum_{i=1}^n x_i = 1$?
6. **Least-squares (10 marks).** Write (Python) code to solve the following least-squares problem over $x \in \mathbb{R}^2$,

$$\text{minimize} \quad \|Ax - b\|_2^2$$

with $A = \begin{bmatrix} 8 & 4 \\ 6 & 6 \\ 5 & 8 \\ 0 & 0 \end{bmatrix}$ and $b = \begin{bmatrix} 2 \\ 0 \\ 2 \\ 2 \end{bmatrix}$. Include the code listing and solution you get from running your code.

7. **Gradients (10 marks).** Compute the gradients $\nabla_x f(x)$ of the following functions:
   (a) $f(x) = a^T x + b$ for $a \in \mathbb{R}^n, b \in \mathbb{R}$
   (b) $f(x) = \frac{1}{2} x^T P x + q^T x + r$ for $P \in \mathbb{R}^{n \times n}, q \in \mathbb{R}^n, r \in \mathbb{R}$
   (c) $f(x) = \frac{1}{2} x^T P x$ for $P = P^T \in \mathbb{R}^{n \times n}$
   (d) $f(x) = (a^T x - b)(c^T x - d)$ for $a, c \in \mathbb{R}^n$ and $b, d \in \mathbb{R}$
   (e) $f(x) = \dfrac{1}{1 + \exp(-g(x))}$ in terms of $f(x)$ and $\nabla_x g(x)$

8. **Trace. (10 marks).** Show that the trace operator is a valid inner product for the vector space of real $m \times n$ matrices. That is, for $X, Y \in \mathbb{R}^{m \times n}$, show that $\langle X, Y \rangle = \mathbf{tr}\left(X^T Y\right)$ satisfies the properties of an inner product.

9. **Anti-symmetric matrix (10 marks).** Consider an arbitrary anti-symmetric matrix $A \in \mathbb{R}^{n \times n}$, i.e., $A^T = -A$. Prove that $x^T A x = 0$ for all $x \in \mathbb{R}^n$. Hence, show that for any matrix $B \in \mathbb{R}^{n \times n}$ we have $\frac{1}{2} x^T (B + B^T) x = x^T B x$.

10. **Maximum entropy (10 marks).** Consider the optimisation problem

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & \mathbf{1}^T x = 1 \end{array}$$

   for $x \in \mathbb{R}^n_{++}$.
   (a) Examine the constraint function and write and expression for x_n in terms of the other x_i.
   (b) Use your expression to replace x_n in the objective function to construct an unconstrained optimisation problem.
   (c) Solve the new (unconstrained) optimisation problem by differentiating the objective and setting it to zero to find the maximum entropy distribution. *Hint: Use symmetry and constraints on x from the original problem.*