The Australian National University Semester 2, 2023
School of Computing Theory Assignment 2 of 4

COMP3670/6670: Introduction to Machine Learning

Release Date. Aug 16th, 2023
Due Date. 11:59pm, Sept 17th, 2023
Maximum credit. 100

Exercise 1 **Orthogonal Projections** (3 + 3 + 3 + 4 + 6 + 3 credits)

Consider the Euclidean vector space $\mathbb{R}^3$ with the dot product. A subspace $U \subset \mathbb{R}^3$ and vector $\mathbf{x} \in \mathbb{R}^3$ are given by:

$$U = \text{span}\left\{ \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ -2 \end{bmatrix} \right\}, \mathbf{x} = \begin{bmatrix} 8 \\ 4 \\ 16 \end{bmatrix}$$

1. Show that $\mathbf{x} \notin U$.
2. Determine the orthogonal projection of $\mathbf{x}$ onto U, denoted $\pi_U(\mathbf{x})$.
3. Determine the distance $d(\mathbf{x}, U) := \min_{\mathbf{y} \in U} \|\mathbf{x} - \mathbf{y}\|$, where $\| \cdot \|$ denotes the Euclidean norm.
4. Use Gram-Schmidt orthogonalization to transform the matrix $\mathbf{A} = \begin{bmatrix} -1 & 2 \\ 1 & -1 \\ 1 & -2 \end{bmatrix}$ into a matrix $\mathbf{B}$ with orthonormal columns.
5. Let $\mathbf{Q} \in \mathbb{R}^{m \times n}$ be a matrix with orthonormal columns and $\mathbf{x} \in \mathbb{R}^m$ be an m-dimensional vector. Find the vector $\boldsymbol{\theta}$ that minimizes $||\mathbf{x} - \mathbf{Q}\boldsymbol{\theta}||^2 + \lambda||\boldsymbol{\theta}||^2$, where λ is a positive real number.
6. Compute the vector $\boldsymbol{\theta}$ for the matrix $\mathbf{B}$ and $\lambda = 10$.

Exercise 2 **Vector calculus practices** (6 + 8 + 8 credits)

Compute the following gradients over $\mathbf{x}$ or $\mathbf{X}$. Represent the result in numerator layout. **Note that you are only allowed to use the rules demonstrated in the lecture**. Show each step clearly.

1. $\dfrac{\partial \mathbf{x}^T \mathbf{ABC}\, \mathbf{x}}{\partial \mathbf{x}}$
2. $\dfrac{\partial (\mathbf{Bx} + \mathbf{b})^T \mathbf{C}(\mathbf{Dx} + \mathbf{d})}{\partial \mathbf{x}}$
3. $\dfrac{\partial\, \mathbf{tr}(\mathbf{X}^2)}{\partial \mathbf{X}}$

Exercise 3 **Concavity of a function** (8 + 10 + 10 credits)

A function $f : \mathbb{R}^n \to \mathbb{R}$ with a convex domain is called a **concave** function if and only if its Hessian $\mathbf{H} = \frac{\partial^2 f}{\partial \mathbf{x}^2}$ is negative semidefinite. Consider the following function:

$$f(\mathbf{x}) = \left(\sum_{i=1}^{n} x_i^p \right)^{1/p}$$

with convex domain $\mathbf{dom}(f) = \mathbb{R}^n_{++}$ (n-dim strictly elementwise positive vectors), and $p < 1, p \neq 0$.

1. Evaluate the elementwise second order derivatives $\frac{\partial^2 f}{\partial x_i x_j}$ for arbitrary integer $i, j \in [1, n]$.

2. Denote the elementwise power of a vector $\mathbf{a} \in \mathbb{R}^n_{++}$ to a real number t as $\mathbf{a}^t = \begin{bmatrix} {a_1}^t & {a_2}^t & \cdots & {a_n}^t \end{bmatrix}^T$. Also, the $\mathbf{diag}(\cdot)$ function returns the diagonal matrix with diagonal values input as a vector. Prove that

$$\mathbf{H} = (1-p)f(\mathbf{x})^{1-2p} \cdot \left(\mathbf{x}^{p-1} \cdot \mathbf{x}^{p-1^T} - f(\mathbf{x})^p \cdot \mathbf{diag}\left(\mathbf{x}^{p-2}\right)\right)$$

3. Prove $\mathbf{H}$ is negative semidefinite, hence f is concave since it has a convex domain.

Exercise 4 Expectations with respect to a Gaussian distribution (10+10+8 credits)

A common objective function in modern machine learning is the variational free-energy,

$$\mathcal{F}(q(\theta)) = \int \mathrm{d}\theta q(\theta) \log \frac{q(\theta)}{p(\theta)p(y|\theta, x)} = \int \mathrm{d}\theta q(\theta)[\log q(\theta) - \log p(\theta) - \log p(y|\theta, x)]. \tag{1}$$

Consider a simplified setting in which

$$p(\theta) = \mathcal{N}(\theta; 0, 1), \tag{2}$$
$$p(y|\theta, x) = \mathcal{N}(y; \theta x, \sigma_n^2), \tag{3}$$
$$q(\theta) = \mathcal{N}(\theta; \mu, \sigma^2), \tag{4}$$

where $\mathcal{N}(x; m, v)$ means x is a univariate Gaussian random variable with mean m and variance v.

1. Compute $\mathcal{F}$.

2. Find the gradients $\frac{\partial}{\partial \mu}\mathcal{F}$ and $\frac{\partial}{\partial \sigma}\mathcal{F}$.

3. Set these gradients to zero and solve for μ and σ in terms of y, x and σ_n.