Introduction to Machine Learning
Midterm

1. The *Vandermonde matrix* is

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{d-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{d-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^{d-1} \end{pmatrix}, \quad n, d \in \mathbb{N}.$$

   (a) Calculate the *Vandermonde determinant* $|\boldsymbol{X}|$ if $n = d = 1, 2, 3$. (**1 point**)

   (b) Find the formula for the Vandermonde determinant $|\boldsymbol{X}|$ in general case ($n = d$). Justify your answer. (**2 points**)

   (c) Prove that the matrix $\boldsymbol{X}$ has full column rank if $n \geqslant d$ and all x_i are different. (**1.5 points**)

2. Let $\boldsymbol{A} = \sum\limits_{i=1}^{k} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\mathsf{T}$, where the system of vectors $\boldsymbol{u}_i \in \mathbb{R}^n$ is orthonormal, $1 \leqslant k \leqslant n$, $\lambda_i \neq \lambda_j$ if $i \neq j$.

   (a) Find the shape of $\boldsymbol{A}$ and prove that it is symmetric. (**1 points**)

   (b) Find $\mathrm{tr}(\boldsymbol{A})$. (**1 points**)

   (c) Find all eigenvalues of $\boldsymbol{A}$. Justify your answer. (**2 point**)

3. Let $\boldsymbol{U} \in \mathbb{R}^{n \times n}$ be an orthogonal matirix. Find its spectral norm $\|\boldsymbol{U}\|_2$, Frobenius norm $\|\boldsymbol{U}\|_F$ and condition number $\kappa(\boldsymbol{U})$. Justify your answers. (**2 points**)

4. Let $\boldsymbol{X} \in \mathbb{R}^{m \times n}$ and $f(\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{X}\boldsymbol{B}\boldsymbol{X}^\mathsf{T})$. Find $\nabla f(\boldsymbol{X})$. Which shape must $\boldsymbol{B}$ have? (**2 points**)

5. *Pareto distribution* with parameters $x_m > 0$ and $\alpha > 0$ has cdf

$$F_\xi(x) = 1 - \Big(\frac{x_m}{x}\Big)^\alpha, \quad x \geqslant x_m.$$

   (a) Find $\mathbb{E}\xi$. For which α does it exist? (**1 point**)

   (b) Find $\mathbb{V}\xi$. For which α does it exist? (**1.5 points**)

   (c) Find $\mathbb{H}[\xi]$. What is the limit of the entropy if $\alpha \to +\infty$? $\alpha \to +0$? (**2 points**)

6. Let $X_1, \dots, X_n$ be an i.i.d sample from $U[0, 2\theta]$. Consider the following estimation of θ:

$$\widehat{\theta} = \frac{X_{(1)} + X_{(n)}}{2} = \frac{1}{2}\min\{X_1, \dots, X_n\} + \frac{1}{2}\max\{X_1, \dots, X_n\}.$$

   (a) Find the bias-variance decomposition of this estimation. (**2 points**)

   (b) Is this estimation unbiased? asymptotically unbiased? consistent? Justify your answer. (**2 points**)