HDDA - List of theoretical questions for the Midterm 2024

1. *Singular Value Decomposition.*
   - Q1: Prove the existence of the SVD for any matrix $A \in \mathbb{R}^{m \times n}$.
   - Q2: State and prove the Eckart-Young-Mirsky Theorem for the choice $\|.\| = \|.\|_2$, that is the spectral norm. What is therefore the expression for A_k, the best rank-k approximation of an input matrix $A \in \mathbb{R}^{m \times n}$.
2. *Linear Regression.*
   - Q3: Explain the process of solving linear least-squares regression, i.e. how to derive the normal equations and how to solve them (where possible).
   - Q4: Show that general least squares problems can be solved using SVD. Explain how we can simply find the optimal θ with the minimal Euclidean norm $\|\theta||_2$.
   - Q5: When using an ℓ_1-norm for the residual $A\theta - b$, state the optimization problem, and reformulate it as a linear program in standard form.
   - Q6: Show that least-squares linear regression, i.e. $f(\theta) = \|A\theta - b\|_2^2$, is a maximum likelihood estimation in disguise.
3. *Principal Component Analysis.*
   - Q7: Build the Problem formulation behind PCA for general k. (you can start for $k = 1$ and then generalize it for any k).
   - Q8: Demonstrate how to obtain the first principal component of a matrix $X \in \mathbb{R}^{m \times n}$ (where $m, n \in \mathbb{N}_0$, $, m > n > 2$), representing m data points in $\mathbb{R}^n$. Based on this result, PCA boils down to calculating which vectors of which matrix?
4. *Shades of Regression.*
   - Q9: Gradient descent for linear regressions with differentiable objective functions f:
     - Define what is a differentiable L-smooth function $f : \mathbb{R}^n \to \mathbb{R}$
     - Write the pseudo-code for the gradient descent (GD) method. Explain the main difficulties encountered when using the GD method.
     - Explain what is the full-relaxation strategy for computing the step size α_k.
     - What is the advantage of the GD method over direct methods when it comes to solving optimization problems for linear regressions?
     - When the function f is L-smooth, explain how to compute the expression of the *optimal constant* step size.
     - What is the value for L when we consider the least squares linear problem, that is for $f(\theta) = \frac{1}{2}\|A\theta - b\|_2^2$?
   - Q10: Logistic regression:
     - Define the logistic regression model (which h-hypothesis?).

- Explain how the θ parameters can be estimated using the maximum likelihood estimation method.
- Write the pseudo-code for the stochastic gradient method to solve the logistic regression problem.
- Compare the expression of the stochastic gradients obtained when solving the least squares linear regression and logistic regression problems. What do you observe ?

5. *Clustering.*

- Q11: Define what could be a "good" cluster, define the mean and the variance of clusters. What is the challenge with such definitions ?
  Formulate the optimization problem behind what we call the "k-Means Clustering" problem.
- Q12: Write the pseudo-code for Lloyd's algorithm. Explain the steps. Does this algorithm guarantee to find the optimal solution to the k-Means Clustering problem? If not, why not (is the k-Means Clustering problem easy to solve?)?
- Q13: Why Does Lloyd's Algorithm Terminate in Finite time?