Subject Code: KAI071

PAPER ID-310304

Roll No: | | | | | | | | | | | | | |

BTECH
(SEM VII) THEORY EXAMINATION 2023-24
OPTIMIZATION IN MACHINE LEARNING

TIME: 3 HRS **M.MARKS: 100**

Note: 1. Attempt all Sections. If require any missing data; then choose suitably.

SECTION A

1. Attempt *all* questions in brief. 2 x 10 = 20

Q no.	Question	Marks
a.	How does convexity play a crucial role in optimization problems?	2
b.	Illustrate the application of convex optimization in real-world scenarios.	2
c.	Explain Nesterov's application in convex optimization.	2
d.	Investigate Moreau–Yosida regularization.	2
e.	Explain the regularization process.	2
f.	Explain the concept of dual decomposition in optimization	2
g.	How Douglas–Rachford splitting addresses challenges?	2
h.	How do optimization algorithms strategically navigate saddle points?	2
i.	What are the implications for convergence and optimization efficiency?	2
j.	How does the structure of the optimization landscape impact the choice of optimization algorithms?	2

SECTION B

2. Attempt any *three* of the following: 10 x 3 = 30

a.	Compare and contrast linear programming, second-order cone programming, and semidefinite programming. Provide real-world examples where each type of convex program is applicable.	10
b.	Explain the concept of duality in convex optimization. How does duality provide insights into the primal and dual aspects of an optimization problem? Discuss the relationship between the primal and dual solutions.	10
c.	Compare and contrast mirror descent and traditional gradient descent methods. Highlight the advantages and disadvantages of mirror descent, and provide a real-world example where mirror descent might outperform other optimization algorithms.	10
d.	Compare the convergence properties of Augmented Lagrangian methods and ADMM. Discuss scenarios where one method might be preferred over the other based on the nature of the optimization problem.	10
e.	Compare and contrast Polyak–Juditsky averaging with other methods used in stochastic gradient descent. How does it contribute to the convergence and stability of optimization algorithms, especially in the context of deep learning?	10

SECTION C

3. Attempt any *one* part of the following: 10 x 1 = 10

a.	Describe the Karush-Kuhn-Tucker (KKT) conditions in the context of convex optimization. How do these conditions characterize optimality in convex programs? Provide examples to illustrate the application of KKT conditions.	10

Subject Code: KAI071

PAPER ID-310304

Roll No:

BTECH
(SEM VII) THEORY EXAMINATION 2023-24
OPTIMIZATION IN MACHINE LEARNING

TIME: 3 HRS **M.MARKS: 100**

b.	Provide detailed explanations and examples of different convex programs. Highlight the specific mathematical formulations and problem structures for each type of convex program.	10

4. Attempt any *one* part of the following: 10 x 1 = 10

a.	Elaborate on the Frank–Wolfe method and its applications in constrained optimization. Discuss how the method addresses challenges posed by large-scale optimization problems and provide a scenario where it is particularly effective.	10
b.	Explore the concept of Ordinary Differential Equations interpretations in the context of optimization algorithms.	10

5. Attempt any *one* part of the following: 10 x 1 = 10

a.	Discuss the role of dual methods in optimization and their applications in solving primal-dual problems. Provide an example illustrating the use of dual methods and explain their advantages over primal methods in certain scenarios.	10
b.	Provide an in-depth overview of proximal gradient methods and their role in handling non-smooth and non-convex optimization problems.	10

6. Attempt any *one* part of the following: 10 x 1 = 10

a.	Compare and contrast the Alternating Direction Method of Multipliers with other optimization algorithms, highlighting its strengths and weaknesses.	10
b.	Describe the Douglas–Rachford splitting algorithm and its application in solving convex optimization problems. Discuss situations where this algorithm is particularly effective and the conditions under which it converges.	10

7. Attempt any *one* part of the following: 10 x 1 = 10

a.	Elaborate on the application of Langevin dynamics in Bayesian inference. How does it relate to escaping saddle points and contribute to efficient sampling in high-dimensional spaces, particularly in the context of deep learning models?	10
b.	Examine the Bayesian interpretation of Langevin dynamics. How does Langevin dynamics connect to the posterior distribution in Bayesian models, and what are the implications for exploring the posterior space in the context of probabilistic modeling?	10