**Department of Electrical Engineering**
**Indian Institute of Technology Kharagpur**

---

**Machine Learning for Signal Processing Laboratory**

(EE69210)

Spring, 2022-23

---

## Experiment 5: Regression and Classification

---

**Grading Rubric**

| | Tick the best applicable per row | | | Points |
|---|---|---|---|---|
| | Below Expectation | Lacking in Some | Meets all Expectation | |
| Completeness of the report | | | | |
| Organization of the report (5 pts) | | | | |
| Quality of figures (5 pts) | | | | |
| SVM 1 (20 pts) | | | | |
| SVM 2 (20 pts) | | | | |
| Regression 1 (40 pts) | | | | |
| Regression 2 (20 pts) | | | | |
| | | | | TOTAL (100 pts) |

# 1 Code of Conduct

Students are expected to behave ethically both in and out of the lab. Unethical behaviour includes, but is not limited to, the following:

- Possession of another person's laboratory solutions from the current or previous years.
- Reference to, or use of another person's laboratory solutions from the current or previous years.
- Submission of work that is not done by your laboratory group.
- Allowing another person to copy your laboratory solutions or work.
- Cheating on quizzes.

The rules of laboratory ethics are designed to facilitate these goals. We emphasize that laboratory TAs are available to help the student both understand the basic concepts and answer the questions being asked in the laboratory exercises. By performing the laboratories independently, students will likely learn more and improve their performance in the course as a whole. Please note that it is the responsibility of the student to make sure that the content of their graded laboratories is not distributed to other students. If there is any question as to whether a given action might be considered unethical, please see the professor or the TA before you engage in such actions.

# 2 Linear Classification using Support Vector Machine

Consider the support vector machine (SVM) classification problem discussed in class. Labeled dataset $\{x^i, y^i\}_{i\in[N]}$ where each $x^i \in \mathbf{R}^n$ is associated with a label $y^i \in \{1, -1\}$ such that $y^i = 1$ if $x^i \in A$ and $y^i = -1$ if $x^i \in B$ is given. Consider the following classification problem:

$$\begin{aligned} \min_{w\in\mathbb{R}^n, b\in\mathbb{R}} \quad & \frac{1}{2}||w||_2^2 \\ \text{s.t.} \quad & 1 - y^i(w^\top x^i + b) \le 0, \qquad \forall i \in [N]. \end{aligned}$$

Using a suitable convex optimization solver, answer the following questions.

1. Solve the primal problem and determine the optimal weights $w$ and $b$.
2. Formulate and solve the dual problem and determine the optimal dual solution.
3. Given the optimal dual solution $\lambda^*$, find the optimal primal solution $w^*$ and $b^*$.
4. Plot the data points as well as the classifier with the data points associated with the support vector highlighted. Use different colors to denote data points that belong to the different sets $A$ and $B$.

# 3 Nonlinear Classification using Kernels

A dataset containing 100 samples in the form $\{x^i, y^i\}_{i\in[N]}$ where each $x^i \in \mathbb{R}^4$ and $y^i \in \{1, -1\}$ is given. Solve the dual of the SVM problem using Gaussian Kernel to find the optimal dual solution. Create a script which will take as input some $x \in \mathbb{R}^4$ and determine its label. Compute the label of each data point in the data set, compare with the given label and determine how many samples are incorrectly classified.

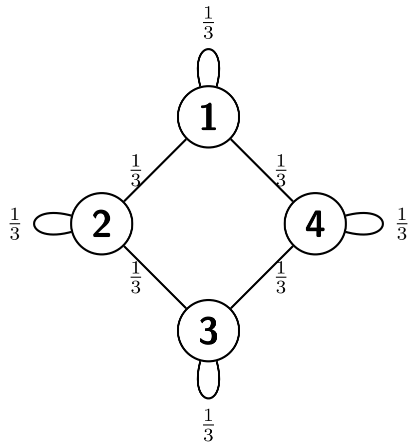


Figure 1: Network Topology for the Distributed Least Squares Problem

## 4 Least Squares Regression

A dataset containing 100 samples in the form $\{x^i, y^i\}_{i\in[N]}$ where each $x^i \in \mathbb{R}^3$ and $y^i \in \mathbb{R}$ is given. The output $y$ is a polynomial of degree at most 2 of the input $x$. Answer the following questions.

1. Define a suitable feature map $\phi(x)$ which maps $x \in \mathbb{R}^3$ to entries of a polynomial of degree 2 in $x$. What is the dimension of $\phi(x)$?
2. Formulate a least squares problem to determine the coefficients of this polynomial. Clearly state the decision variable $w$, its dimension, and the cost function.
3. Compute the gradient of this cost function with respect to the decision variable $w$.
4. Find the optimal weights $w^\star$ using a suitable solver.
5. Compute the error vector $y^i - \phi(x^i)^\top w^\star$ and plot its histogram.
6. Solve the above problem using Gradient Descent, Accelerated Gradient Descent and Stochastic Gradient Descent for 1000 steps. Plot the value of the cost function computed at each iteration and the error $||w_t - w^\star||$ vs. number of interations for all three algorithms.

Hint: For a quadratic function $x^\top Ax$, its smoothness parameter is the largest eigenvalue of $A$ and the strong convexity parameter is the smallest eigenvalue of $A$.

## 5 Distributed Least Squares Regression

Consider a distributed version of the above least squares problem where there are four agents $\{1, 2, 3, 4\}$. The communication topology among the agents and the weights between any two neighboring agents are shown in Figure 1. Suppose the above dataset is held by the agents as follows: the first 20 data points are with agent 1, the next 25 data points are with agent 2, the next 30 data points are with agent 3 and the final 25 data points are with agent 4. Use distributed gradient descent for 100000 steps to solve the regression problem.

1. Initialize a local solution $w_i^0$ for each agent $i$ randomly.

2. Determine the weights $a_{ij}$ values for each pair of nodes.
3. Choose step size $\eta_t = \frac{1}{2\beta\sqrt{t}}$ where $\beta$ is the smoothness parameter.
4. Plot $||w_i^t - \bar{w}^t||^2$ for each agent $i$ (in log scale) and show convergence of these quantities to 0.
5. Plot $||w^\star - \bar{w}^t||^2$ and show convergence of this quantity to 0.