Introduction to ML (CS771), Autumn 2023
Indian Institute of Technology Kanpur
Homework Assignment Number 1

QUESTION 1

Student Name: Soham Bharambe
Roll Number: 210264
Date: December 10, 2023

Optimization problem is,

$$\left(\hat{\boldsymbol{w}}_c, \hat{\mathbf{M}}_c\right) = \arg \min_{\boldsymbol{w}_c, \mathbf{M}_c} \sum_{\boldsymbol{x}_n : y_n = c} \frac{1}{N_c} \left(\boldsymbol{x}_n - \boldsymbol{w}_c\right)^\top \mathbf{M}_c \left(\boldsymbol{x}_n - \boldsymbol{w}_c\right) - \log |\mathbf{M}_c|$$

We will be using partial differentiation for finding the optimal solution for both $\hat{\boldsymbol{w}}_c$ and $\hat{\mathbf{M}}_c$.
For $\hat{\boldsymbol{w}}_c$,

$$\sum_{\boldsymbol{x}_n : y_n = c} \frac{\partial}{\partial \boldsymbol{w}_c} \left(\left(\boldsymbol{x}_n - \boldsymbol{w}_c\right)^\top \mathbf{M}_c \left(\boldsymbol{x}_n - \boldsymbol{w}_c\right) \right) = \frac{\partial}{\partial \boldsymbol{w}_c} \left(\log |\mathbf{M}_c|\right)$$

$$\sum_{\boldsymbol{x}_n : y_n = c} \left(-2\mathbf{M}_c \left(\boldsymbol{x}_n - \boldsymbol{w}_c\right)\right) = 0$$

$$-2\mathbf{M}_c \sum_{\boldsymbol{x}_n : y_n = c} \left(\boldsymbol{x}_n - \boldsymbol{w}_c\right) = 0$$

$$\sum_{\boldsymbol{x}_n : y_n = c} \left(\boldsymbol{x}_n - \boldsymbol{w}_c\right) = 0$$

$$\sum_{\boldsymbol{x}_n : y_n = c} \boldsymbol{x}_n = \sum_{\boldsymbol{x}_n : y_n = c} \boldsymbol{w}_c$$

$$\hat{\boldsymbol{w}}_c = \frac{1}{N_c} \sum_{\boldsymbol{x}_n : y_n = c} \boldsymbol{x}_n$$

Similarly for $\hat{\mathbf{M}}_c$,

$$\sum_{\boldsymbol{x}_n : y_n = c} \left(\boldsymbol{x}_n - \boldsymbol{w}_c\right) \left(\boldsymbol{x}_n - \boldsymbol{w}_c\right)^\top = \mathbf{M}_c^{-\top}$$

$$\hat{\mathbf{M}}_c = \left(\sum_{\boldsymbol{x}_n : y_n = c} \left(\boldsymbol{x}_n - \boldsymbol{w}_c\right) \left(\boldsymbol{x}_n - \boldsymbol{w}_c\right)^\top \right)^{-\top}$$

$$\hat{\mathbf{M}}_c = \left(\sum_{\boldsymbol{x}_n : y_n = c} \left(\boldsymbol{x}_n - \boldsymbol{w}_c\right) \left(\boldsymbol{x}_n - \boldsymbol{w}_c\right)^\top \right)^{-1}$$

If $\hat{\mathbf{M}}_c$ is an identity matrix then the optimization equation boils down to,

$$\hat{\boldsymbol{w}}_c = \arg\min_{\boldsymbol{w}_c} \sum_{\boldsymbol{x}_n : y_n = c} \frac{1}{N_c} \left\|\boldsymbol{x}_n - \boldsymbol{w}_c\right\|^2$$

This $\hat{\boldsymbol{w}}_c$ will basically give the best fit for all the $\boldsymbol{x}_n$.

Student Name: Soham Bharambe
Roll Number: 210264
Date: December 10, 2023

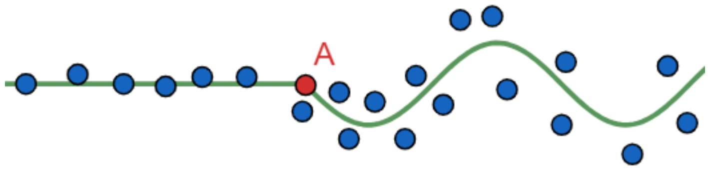
QUESTION 2

With infinite amount of training data, the model should be trained on any test input one can come up with. So the nearest neighbour for that test point would lie on the same coordinates.

Test point will be classified correctly, hence, one-nearest-neighbor algorithm is consistent.

Student Name: Soham Bharambe
Roll Number: 210264
Date: December 10, 2023

QUESTION 3



Suppose given set of points which follow the trend shown in the picture above. Under normal regression one might get a horizontal line as best fit line. What if divide these points at the point A, and then make different regressions for each half? We will get the best fit line as shown in the figure. (I took sine wave but it can also be two different lines with different slopes)

For the split we have to find the point where, if split, the loss function decreases drastically, i.e. the two splits should exhibit somewhat different regression models.

Introduction to ML (CS771), Autumn 2023
Indian Institute of Technology Kanpur
Homework Assignment Number 1

Student Name: Soham Bharambe
Roll Number: 210264
Date: December 10, 2023

QUESTION 4

For the non-regularized linear regression model, solution equation is

$$f\left(\boldsymbol{x}_*\right) = \boldsymbol{x}_*^\top \hat{\boldsymbol{w}}$$

$$\hat{\boldsymbol{w}} = \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

$$f\left(\boldsymbol{x}_*\right) = \boldsymbol{x}_*^\top \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

$$\boldsymbol{w}^\top = \boldsymbol{x}_*^\top \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top$$

$$f\left(\boldsymbol{x}_*\right) = \boldsymbol{w}^\top \boldsymbol{y}$$

$\boldsymbol{w}$ gives us the contribution of each training input to get the prediction of the test output.
The weighted sum over y_i with weights given by the contributions w_i will give us the resultant prediction.

Introduction to ML (CS771), Autumn 2023
Indian Institute of Technology Kanpur
Homework Assignment Number 1

Student Name: Soham Bharambe
Roll Number: 210264
Date: December 10, 2023

QUESTION 5

Loss function for masked input is,

$$L_{mask}\left(\boldsymbol{w}\right) = \sum_{n=1}^{n}\left(y_n - \boldsymbol{w}^\top \tilde{\boldsymbol{x}}_n\right)^2$$

$$L_{mask}\left(\boldsymbol{w}\right) = \sum_{n=1}^{n}\left(y_n - \left(\boldsymbol{w}^\top \boldsymbol{x}_n\right) \circ \boldsymbol{m}_n\right)^2$$

For Expectation of Loss function,

$$\mathbb{E}\left[L_{mask}\left(\boldsymbol{w}\right)\right] = \sum_{n=1}^{n}\mathbb{E}\left[\left(y_n - \left(\boldsymbol{w}^\top \boldsymbol{x}_n\right) \circ \boldsymbol{m}_n\right)^2\right]$$

$$\mathbb{E}\left[L_{mask}\left(\boldsymbol{w}\right)\right] = \sum_{n=1}^{n}\mathbb{E}\left[\left(y_n^2 - 2y_n\left(\left(\boldsymbol{w}^\top \boldsymbol{x}_n\right) \circ \boldsymbol{m}_n\right) + \left(\left(\boldsymbol{w}^\top \boldsymbol{x}_n\right) \circ \boldsymbol{m}_n\right)^2\right)\right]$$

Every element of $\boldsymbol{m}_n$ will be 1 with probability p, and so, expectation for every element of $\boldsymbol{m}_n$ will be p.

$$\mathbb{E}\left[L_{mask}\left(\boldsymbol{w}\right)\right] = \sum_{n=1}^{n}\left(y_n^2 - 2py_n\left(\boldsymbol{w}^\top \boldsymbol{x}_n\right) + p\left(\boldsymbol{w}^\top \boldsymbol{x}_n\right)^2\right)$$

$$\mathbb{E}\left[L_{mask}\left(\boldsymbol{w}\right)\right] = \sum_{n=1}^{n}\left(y_n^2 - 2y_n\left(\boldsymbol{w}^\top \boldsymbol{x}_n\right) + \left(\boldsymbol{w}^\top \boldsymbol{x}_n\right)^2\right) - \sum_{n=1}^{n}\left(2\left(p-1\right)y_n\left(\boldsymbol{w}^\top \boldsymbol{x}_n\right) + \left(1-p\right)\left(\boldsymbol{w}^\top \boldsymbol{x}_n\right)^2\right)$$

$$L_{mask}\left(\boldsymbol{w}\right) = \sum_{n=1}^{n}\left(y_n - \boldsymbol{w}^\top \boldsymbol{x}_n\right)^2 - \sum_{n=1}^{n}\left(2\left(p-1\right)y_n\left(\boldsymbol{w}^\top \boldsymbol{x}_n\right) + \left(1-p\right)\left(\boldsymbol{w}^\top \boldsymbol{x}_n\right)^2\right)$$

$$L_{mask}\left(\boldsymbol{w}\right) = L\left(\boldsymbol{w}\right) + \lambda R\left(\boldsymbol{w}\right)$$

Comparing the above two equations, we get,

$$R\left(\boldsymbol{w}\right) = \sum_{n=1}^{n}\left(2y_n\left(\boldsymbol{w}^\top \boldsymbol{x}_n\right) - \left(\boldsymbol{w}^\top \boldsymbol{x}_n\right)^2\right)$$

$$\lambda = 1 - p$$

Introduction to ML (CS771), Autumn 2023
Indian Institute of Technology Kanpur
Homework Assignment Number 1

Student Name: Soham Bharambe
Roll Number: 210264
Date: December 10, 2023

QUESTION 6

For Method 1 (convex.py)
Accuracy = 0.469

For Method 2 (regress.py)

Lambda	Accuracy
0.01	0.581
0.1	0.595
1	0.674
10	0.733
20	0.717
50	0.651
100	0.565

Table 1: Lambda-Accuracy Table