JAYPEEUNIVERSITYOFINFORMATIONTECHNOLOGY
WAKNAGHAT, P.O.–WAKNAGHAT, TEHSIL– KANDAGHAT,
DISTRICT– SOLAN (H.P.) PIN–173234(INDIA)
(Established by H.P. State Legislature vide Act No. 14 of 2002)

JAYPEE EDUSPHERE
IGNITED MINDS INSPIRED SOULS

18B1WCI634: MACHINE LEARNING

Assignment-1

Submission Date: Feb 7, 2026

Instructions

- Answer all questions.
- Show complete calculations and reasoning.
- Use natural logarithm (log).
- Numerical answers should be rounded to 3 decimal places.

Question 1: Regression with Outliers and Robust Losses

A regression model $h \in \mathscr{H}$ is learned using empirical risk minimization:

$$\hat{h} = \arg\min_{h \in \mathscr{H}} \frac{1}{N} \sum_{i=1}^{N} L\big(h(x_i), y_i\big)$$

Consider the following dataset containing an outlier:

i	True value y_i	Prediction $h(x_i)$
1	2	2.5
2	3	2.8
3	4	4.2
4	5	4.9
5	20	6.0

(a) Mean Squared Error (MSE)

$$L(h(x_i), y_i) = (h(x_i) - y_i)^2$$

- Compute the loss for each data point.
- Compute the overall MSE.
- Identify which data point dominates the MSE.

(b) Root Mean Squared Error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(h(x_i) - y_i)^2}$$

- Compute RMSE.
- Explain why RMSE is preferred for reporting model performance.

(c) Mean Absolute Error (MAE)

$$L(h(x_i), y_i) = |h(x_i) - y_i|$$

- Compute MAE.
- Compare MAE with MSE numerically.

(d) Huber Loss

Huber loss is defined as:

$$L_\delta(h(x_i), y_i) = \begin{cases} \frac{1}{2}(h(x_i) - y_i)^2, & |h(x_i) - y_i| \le \delta \\ \delta|h(x_i) - y_i| - \frac{1}{2}\delta^2, & \text{otherwise} \end{cases}$$

Let $\delta = 1$.

- Compute the Huber loss for each data point.
- Compare its behavior with MSE and MAE.

(e) Conceptual Analysis

- Which loss function is most robust to outliers?
- Which loss would you choose if the data contains frequent extreme values? Justify.

Question 2: Multi-Class Classification and Probabilistic Losses

Consider a 3-class classification problem where the model outputs a probability vector:

$$h(x_i) = (h_1(x_i), h_2(x_i), h_3(x_i))$$

The dataset is given below:

i	True class y_i	Predicted probabilities $h(x_i)$
1	1	(0.7, 0.2, 0.1)
2	2	(0.1, 0.6, 0.3)
3	3	(0.2, 0.2, 0.6)
4	2	(0.4, 0.4, 0.2)

(a) Categorical Cross-Entropy

For one-hot encoded labels:

$$L(h(x_i), y_i) = -\sum_{k=1}^{3} y_{ik} \log h_k(x_i)$$

- Compute the loss for each data point.
- Compute the average cross-entropy loss.

(b) Sparse Categorical Cross-Entropy

$$L(h(x_i), y_i) = -\log h_{y_i}(x_i)$$

- Compute the sparse cross-entropy loss.
- Verify numerically that it matches part (a).

(c) Effect of Confidence

- Identify which sample has the highest loss.
- Explain how prediction confidence affects the loss value.

(d) Cross-Entropy and KL Divergence

Let $P(y \mid x_i)$ be the true distribution and $Q(y \mid x_i) = h(x_i)$.

- Write the expression for cross-entropy $H(P,Q)$.
- Write the expression for $\mathrm{KL}(P \| Q)$.
- Explain why minimizing cross-entropy minimizes KL divergence.

(e) Model Comparison

Suppose a second model produces lower classification accuracy but lower average cross-entropy loss.

- Is this possible? Explain.
- Which model is better from a probabilistic learning perspective?