Road Curve and Straight Section Classification
Project-2 MTH312

1 Group 9 Members:

- Vatsal Rajput - Roll No: 221179
- Sarath Kumar - Roll No: 220754
- Rohit Karwa - Roll No: 220911
- Mukesh Kumar Suthar - Roll No: 231080059

2 Introduction

In this project, we focus on the classification of road segments as either curves or straight sections based on a set of points defined by their X and Y coordinates. The goal is to identify whether a given road section belongs to a curve or a straight segment, using a method that estimates the Radius of Curvature (ROC) at each data point. The ROC is used as a key indicator to classify points, and a threshold is applied to distinguish between curved and straight sections. The method involves preprocessing the data, computing local curvature for each segment. This project utilizes road data from two datasets to assess the model's accuracy and its performance in real-world scenarios.

3 Methodology for Curve Point Identification

In this project, we aim to classify road segments as curves or straight using a dataset of points representing the road in the XY dimension. The methodology employed to achieve this classification is as follows:

1. **Radius of Curvature (ROC) Calculation:** To classify road segments as curves or straight, we first calculate the Radius of Curvature (ROC) for each data point. The ROC is calculated using three non-collinear points along the road. Specifically, for a given data point i, we consider the points at $i-4$, i, and $i+4$ (where i ranges from 5 to $n-4$, with n being the total number of points). These three points define a triangle, and the approximate circumradius of this triangle is calculated to estimate the ROC. The formula used for the approximate of circumradius R of a triangle with side lengths a, b, and c and area A is:

$$R = \frac{a^2b^2 + b^2c^2 + c^2a^2}{4 \cdot A}$$

where a, b, and c are the lengths of the triangle's sides and A is the area of the triangle. If the three points which we are choosing for ith data point come out to be collinear , the area

A of the triangle will become zero. In such cases, we assign a very large value (1e16) to the ROC of that point to handle the division by zero scenarios.

Reason to look for (i+4)th point and (i+4)th point: In our datasets the points on curve are so close that when we calculate ROC using three adjacent points ROC calculation is very unstable. Hence we have taken (i+4)th point and (i+4)th point to see the curve at broader level.

2. **Assigning ROC Values:** For each data point from the 5th to the $n-4$th, the corresponding ROC value is calculated using the above formula. For the first four points (1st to 4th), we assign the ROC value of the 5th point to all of them. Similarly, for the last four points ($n-3$, $n-2$, $n-1$, and n), we assign the ROC value of the $n-4$th point to them.

3. **Log Transformation of ROC Values:** To compress the ROC values and ensure better handling of extreme values, we apply a log transformation. This helps in scaling the ROC values to a more manageable range, which is especially useful for further analysis and classification.

4. **Threshold Definition for Classification:** Once the log-transformed ROC values for all data points are calculated, a threshold is defined to classify points into curves and straight sections. The threshold is calculated as the sum of the minimum ROC value and the standard deviation of the ROC values:

$$\text{Threshold} = \min(ROC) + \text{SD}(ROC)$$

If a point's ROC value is less than the threshold, it is classified as a part of a curve ($Curve = 1$). If the ROC value is greater than or equal to the threshold, the point is classified as part of a straight section ($Curve = 0$).

Reason why we chose threshold in this way: First assume the case where the road has both curves and straight section. Now the point which lies on sharpest curve will have very less ROC but the points which lies on straight line have a constant ROC. We wanted the size of window of predicting a point on curve to depend on the standard deviation of ROC so that when we have points on very sharp curve then our standard deviation of ROC is high and we are able to predict the points as 1(on curve) which are lying on mild curves relative to our sharpest point. And when all points lie on straight line our standard deviation becomes nearly zero and we are able to predict the other points off curve.

5. **Final Output:** The result is a data frame containing the original X, Y coordinates along with the newly calculated **Curve** classification. This classification is represented as a binary value, where 1 indicates a curve section and 0 indicates a straight section of the road.

4 Summary

The methodology for classifying road segments as curves or straight sections relies on calculating the Radius of Curvature (ROC) at each data point. To ensure stability in curvature estimation, we consider three points: the current point at index i along with two additional points at $i - 4$ and $i + 4$. This choice helps mitigate instability in ROC calculations, which occurs when using adjacent points, particularly in densely sampled curved sections. By selecting points farther apart, we obtain a smoother and more reliable estimate of the curvature. Once the ROC values are computed, a log transformation is applied to handle extreme values and improve numerical stability. To classify points as part of a curve or a straight section, a threshold is defined as the sum of the minimum ROC value and the standard deviation of ROC values. This approach ensures that the classification adapts dynamically to the dataset. If a road segment contains both sharp curves and straight sections, the threshold adjusts based on the variation in ROC values, allowing for the identification of both sharp and mild curves. In cases where the entire road is straight, the standard deviation of ROC values is nearly zero, preventing misclassification of straight points as curved. This adaptive thresholding ensures robustness across different road geometries.