PRML Report

**Authors:** CS22B2009, CS22B2045, CS22B2048

# Introduction

The primary goal of this report is to analyze the relationship between image- based features and BMI values through SVM regression while leveraging HOG feature extraction for feature encoding from images and to plot the offense graph and to Predct the gender based on facial features using KNN.

# Meta Data

* Original Data Points: 68,493
* Data After Preprocessing: 60,002
* Training Data: 48,002
* Test Data: 12,000
* Number of Features per image : 6
* Number of males: 56,353
* Number of females: 3,649

# Feature Extraction with HOG

## Why HOG?

HOG (Histogram of Oriented Gradients) features were selected for their effec- tiveness in object recognition and analysis tasks. These features are robust against minor variations in lighting, viewpoint, and deformations. HOG cap- tures the edge information, shape, and texture of an object—important factors when dealing with image analysis tasks like BMI prediction using images.

## Feature Extraction Process

The process followed for image feature extraction was as follows:

### Preprocessing Images:

* + - * Resize all images to a standard size of **128 × 128 pixels**.
      * Convert the images into grayscale to remove color dependencies and simplify feature extraction.
    1. **Extracting HOG Features:** Extract HOG features using a cell size of

**8 × 8** and a block size of **2 × 2**.

### Integrating Features:

* + - * Combine HOG features extracted from both **front and side images**.
      * Store these features in a CSV format for further analysis.

# BMI Prediction with SVM Regression

## Why SVM Regression?

Support Vector Machine (SVM) regression was implemented for BMI prediction because of its ability to handle non-linear relationships between features and target variables. SVM regression uses kernel functions to map data into higher- dimensional spaces where linear separability becomes possible. SVM models are robust, generalizable, and efficient for small to medium-sized datasets like those involved here.

## Preprocessing

Before feeding features into the SVM model, preprocessing was performed with the following steps:

### Conversion of Raw Data into Numerical Arrays:

Convert extracted feature strings into numerical **NumPy arrays**.

### Feature Scaling:

Normalize features using **StandardScaler** from sklearn.preprocessing. This ensures all features are on the same scale, improving SVM model con- vergence and accuracy.

# Gender prediction with KNN

## Why KNN?

K-Nearest Neighbors (KNN) was selected for gender prediction because of its simplicity, effectiveness in binary classification tasks, and its ability to classify data points based on proximity to their nearest neighbors. makes no assumptions about the data distribution, making it robust and suitable for feature-based classification tasks like gender prediction.

## Data Preprocessing

Before implementing the KNN model, preprocessing was performed with the following steps:

### Feature Scaling:

Normalize the extracted HOG features to ensure that all features con- tribute equally during distance calculation.

### Label Encoding:

Encode gender labels into binary values, i.e., **1 for male** and **0 for fe- male**, to prepare them for classification with KNN.

## KNN Model Implementation

The KNN model was implemented with k=50 to ensure that the nearest 50 neighbors are considered during classification. The classification performance was evaluated based on metrics

# Offense Plot

* Extracted the offenses done by people from the sentencing.csv file.
* Counted number of times each offense occurred i.e number of people com- mitted the offense.
* Plotted it using matplotlib library.

# Results

### BMI prediction:

**–** MAE: 3.95

**–** MSE: 27.79

* + R square: -0.01
  + pearson coefficient: 0.10

### Gender prediction:

* + MAE: 0.11
  + MSE: 0.08
  + R square: -0.39
  + pearson coefficient: 0.08