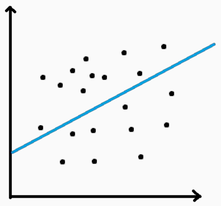
Predicting Heart Attack History in Men Through The Usage of Support Vector Machines

Aakash Krishna | 04/29/2025

Technical Background

Support Vector Machines (SVMs for short) are supervised machine learning models that compute an optimal “hyperplane”, which you can think of as a line that separates individual data points that are in different classes. There are three kinds of SVMs we will concern ourselves with:

Linear



Classes are separated by a straight line/flat plane.

$$K(x_i, x_{i'}) = \sum_{j=1}^{p} x_{ij} x_{i'j}$$

They are best when data is linearly separable.

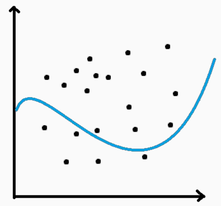
Radial

Classes are judged as similar based on the distance between points.

$$K(x_i, x_{i'}) = \left(1 + \sum_{j=1}^{p} x_{ij} x_{i'j}\right)^d$$

They are best when data is highly non-linear.

Polynomial



Classes are separated by curved boundaries up to a specified degree.

$$K(x_i, x_{i'}) = \exp\left(-\gamma \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2\right)$$

They are best when data is slightly non-linear.

Methods

Our target variable is HEARTATTEV, a measure of if the patient has ever reported a heart attack. In order to predict this, we use the following features (red variable name means said feature was excluded from radial model, blue variable name means an exclusion from the polynomial model, normal text color indicates it was used in all three models):

Tuning Parameters	
Cost (c)	Inversely proportional to margin size, “budget” for points on wrong margin side
Gamma (γ)	Inversely proportional to sphere of influence where similarity of points matters
Degree (d)	Number of curves in line

Variable Name	Measure
HEIGHT	Height in inches without shoes
BMICALC	Calculated Body Mass Index
MOD10DMIN	Average duration of moderate activity per day
JUICEMNO	Number of times consuming 100% fruit juice in last month
SALSAMNO	Number of times consuming salsa in last month
TOMSAUCEMNO	Number of times consuming tomato sauce in last month
SPORDRMNO	Number of times consuming sports drinks in last month
FRTDRINKMNO	Number of times consuming sugary fruit drinks in last month
COFETEAMNO	Number of times consuming coffee/tea in last month
HRSLEEP	Usual hours of sleep per day
ALCDAYSM	A variable created from ALCDAYSYR in the dataset, to record number of days alcohol is consumed per month on average.
HTATK	Factor created from HEARTATTEV to make predicting HEARTATTEV easier.

Results