2 Methodology

2.1 Nadarayan-Watson Kernel estimator

Nadarayan-Watson kernel estimator is an important nonparametric kernel estimator of a regression function which is achieved by using a fixed bandwidth. The NW method is a method for estimating unknown parameters of a regression function and is suitable for the situation where the data comes from a joint probability distribution function, f (x, y). The non-parametric regression model for NW method is

$$Y_i = M(X_i) + e_i \tag{1}$$

for (i = 1,……,n)

Where m (.) is unknown, $\mathrm{Y_i}$ is the sum of the regression function value $\mathrm{m(X_i)}$ for $\mathrm{X_i}$ and predicted error value is zero.

f (x, y) is the joint density function of (X, Y) and f(x) is the marginal density function of f(x). An estimation of this regression function can be taken as

$$M_n(x) = \int \frac{\mathrm{y}\,\hat{\mathrm{f}}(\mathrm{x,y})}{\mathrm{f(x)}}\, dy \tag{2}$$

Using kernel estimation k(.), we have nonparametric kernel estimation of regression.

$$\hat{f}(x,y) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_1 h_2} K^2\left(\frac{x-X_i}{h_1}, \frac{y-Y_i}{h_2}\right) \tag{3}$$

Using bivariate kernel function and substituting different smoothing parameters (h1, h2) with single fixed smoothing parameter (h) we get the NW kernel estimator of regression.

$$\widehat{M}_n(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right).Y_i}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)} \tag{4}$$

Smoothing level of estimation also known as bandwidth (h) plays a very important role in the performance of the kernel estimators. The commonly used methods for deciding h are cross-validation, penalized functions, and bootstrap. (Demir & Toktamiş, 2010)

2.2 Support Vector Machines

The idea of SVM was established in 1958 by Rosenblatt. SVM is one of the popular and efficient classification and regression methods which separates hyperplane with maximum margin in two-dimensional space which are linearly separable. The one that maximizes the distance to the closest data points from both classes is called hyperplane with maximum margin. However, for a non-linear data SVM finds it difficult to classify the data. That’s where we can use the Kernel Trick. A Kernel trick is a method where a non-linear data is projected onto a higher dimension space to make it easier to classify the data where it could be linearly divided by a plane (Karatzoglou, Meyer, & Hornik, 2006). SVMs use an implicit mapping Φ of the input data into a high-dimensional feature space. The common classification model is

$$f(x) = sign(\langle w,\phi(x)\rangle + b) \quad (5)$$

Where *b* is the bias of the model.

Solving the subsequent constrained optimization and creating decision boundary, we get

$$t(w,\varepsilon) = \frac{1}{2} \| w \|^2 + \frac{C}{m}\sum_{i=1}^{m} \varepsilon_i \quad (6)$$

$$\textit{Subject to } y_i(\langle w,\phi(x)\rangle + b) \geq 1 - \varepsilon_i$$

Where (i = 1,…..m), $\varepsilon_i \geq 0$

As it is constrained optimization problem it needs Lagrangian formula or Lagrangian multipliers. Firstly, we need to find decision boundary for linearly separable plane and then solve with Lagrange multipliers. We get following equation,

$$maximize\ f(c_1 \ldots c_n) = \sum_{i=1}^{n} c_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_1\, c_i\, k(x_i, x_j) y_j c_j$$

$$\text{subject } \sum_{i=1}^{n} c_i\, y_i = 0 \text{ and } 0 \leq c_i \leq \frac{1}{2n\lambda} \text{ for all } i.$$

As there are different types of kernel function for different type of data, picking right kernel and parameters can be the challenge for SVMs.

Types of Kernel function

There are five main types of kernels which are listed below with equations.

1. Linear Kernel Function

The linear kernel is the basic and useful when dealing with large sparse data vectors. Linear Kernel is the inner product of two vectors. The equation is:

$$K(xi, xj) = xi.xj + c \quad (8)$$

Where c is the constant.

2. Polynomial Kernel Function

The polynomial kernel is commonly used function with SVMs. It represents kernels with a degree of more than one. It helps on learning of non-linear models and suitable for image processing. (Vidvan, 2020). There are two types of polynomials:

Homogenous Polynomial Kernel Function

$$K(xi, xj) = (xi.xj)^{\mathrm{d}} \quad (9)$$

where d is the degree of the polynomial.

Inhomogeneous Polynomial Kernel Function

$$K(xi, xj) = (xi.xj + c)^{\mathrm{d}} \quad (10)$$

where c is a constant and d is degree of polynomial.

3. Gaussian RBF Kernel Function

Gaussian RBF is used when there is no prior knowledge about the data.

$$K(xi, xj) = exp(-\gamma||xi - xj||)^2 \quad (11)$$

Where γ is parameter to optimize.

4. Sigmoid Kernel Function

This is mainly used in neural networks. We can show this as:

$$K(xi, xj) = tanh(\alpha xay + c) \quad (12)$$

Where h and c are parameters to optimize.

Karatzoglou, A., Meyer, D., & Hornik, K. (2006). Support Vector Machines in*R*. *Journal of Statistical Software*, *15*(9). https://doi.org/10.18637/jss.v015.i09

Vidvan, T. (2020, March 13). SVM Kernel Functions - 'SVM knowledge. Retrieved from TechVidvan website: https://techvidvan.com/tutorials/svm-kernel-functions/