Support Vector Machine Modeling on Tick Data

Fitting and Evaluating SVM Models

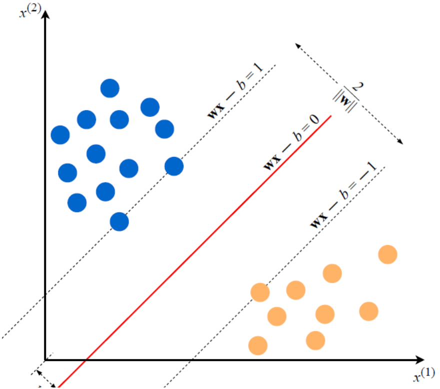
Radha Pendyala

REFINITIV™

Support Vector Machines

- Real world datasets have features that are usually not *linearly separable*
- One might have to create additional *features* from the existing features for classification or regression tasks
- SVM is one of the popular algorithms that helps you map the features to higher dimensional feature space with out additional *computational burden*.

FIT A LARGE MARGIN BETWEEN THE POINTS



Support Vector Machines

- The higher the $||\mathbf{w}||$, the higher the separation. One would ideally want a large separation. However it might not lead to good classification
- Key idea is that there are support vectors that are only dependent on a *few training samples* out of the entire dataset
- SVM is robust to *outliers*
- Linear SVM Model for Classification(hard-margin SVM) for linearly separable cases

$$\min \frac{1}{2}||w||^2$$
$$\mathbf{wx}_i - b \geq 1,\ y_i = +1$$
$$\mathbf{wx}_i - b \leq -11,\ y_i = -1$$

  - minimize the norm $\equiv$ maximizing the margin of the classifier
- Linear SVM Model for Classification(soft-margin) using *Hinge-Loss function*

$$\min C||w||^2 + \frac{1}{N}\sum_{i=1}^{N}\max(0, 1 - y_i(\mathbf{wx}_i - b))$$

High value of C leads to highest margin and Low value of C leads to lesser misclassification

Support Vector Machines

- Linear SVM Model for Regression
  - Loss function structure

$$\frac{1}{N}\sum_{i=1}^{N}\max(0, |y_i - (\mathbf{w}\mathbf{x}_i + b)| - \epsilon)$$

  - ϵ is the hyperparameter
  - fit as many points as possible in the 2ϵ width margin from the hyperplane
- Non Linear SVM Model for Classification

$$\max \sum_{i=1}^{N}\lambda_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{N} y_i\lambda_i\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)y_k\lambda_k$$

$$\sum_{i=1}^{N}\lambda_i y_i = 0$$

$$\lambda_i \geq 0 \; \forall i \in \{1, \dots, N\}$$

  - λ_i are *Lagrangian Multipliers*
  - $\mathbf{K}$ is the *Kernel* that enters in to all the computations

REFINITIV

Support Vector Machines

- Popular Kernel functions
  - Linear Kernel

$$\mathbf{K}(\mathbf{x}, \mathbf{y}) = (r + \mathbf{x}^T \mathbf{y})$$

  - Polynomial Kernel

$$\mathbf{K}(\mathbf{x}, \mathbf{y}) = (r + \mathbf{x}^T \mathbf{y})^d$$

  - Radial Basis Function(RBF)

$$\mathbf{K}(\mathbf{x}, \mathbf{y}) = \exp(-\gamma(\mathbf{x} - \mathbf{y})^2)$$

  - Sigmoid Kernel

$$\mathbf{K}(\mathbf{x}, \mathbf{y}) = \tanh(\gamma \mathbf{x}^T \mathbf{y} + r)$$

- Each of the kernels have different hyperparameters and hence the optimization algorithm has to be coded accordingly
- Off-the shelf libraries provide relevant arguments to invoke specific kernels

SVM on Tick Data

- Tick Data considered for one of the stocks trading in the Hong Kong market
- Main idea is to compare the price predictions of SVM with a random walk and check whether it beats random walk model in terms of error rates
- Dataset spans **5** months of Tick data for the stocks trading on HongKong Stock exchange
- Technical Indicators(**35**) have been used as features for SVM Model
  - Momentum Indicators
  - Volatility Indicators
  - Trend Indicators
  - Simple Average and Weighted Average
- SVM model fitted to 1 min prices and Kernels used are
  - Linear Kernel
  - Radial Basis Kernel
  - Polynomial Kernel
- Grid Search combined with Cross validation has been used to estimate hyperparameters
- Root-mean-squared error(**RMSE**) and Mean Absolute Percentage Error(**MAPE**) are computed for SVM model as well as Random walk model

Technical Indicators used as features

- Awesome Oscillator
- Relative Strength Index
- Stochastic Oscillator
- Stochastic Oscillator Signal
- True strength index
- Ultimate Oscillator
- Williams %R
- Average True Range
- Bollinger High Band
- Bollinger High Band Indicator
- Bollinger Lower Band
- Bollinger Low Band Indicator
- Donchian High Band
- Donchian High Band Indicator
- Donchian Low Band
- Donchian Low Band Indicator
- Keltner channel High Band Indicator
- Keltner channel Low Band
- Keltner channel Low Band Indicator
- Aroon Indicator Up
- Aroon Indicator Down
- Detrended Price Oscillator
- Ichimoku Above
- Ichimoku Below
- KST Oscillator
- KST Oscillator Signal
- MACD
- MACD Difference
- MACD Signal
- Mass Index
- Trix
- Vortex Indicator Pos and Neg

Hyperparameter Tuning and Performance metrics

- Hyperparameters C plays in controlling the margin width amongst classifiers and regressors. Higher the C, the more the risk of overfitting
- Hyperparameter d plays a role in a role in Polynomial SVM Kernel. Increasing the degree of the polynomial. The higher the degree, the more the risk of overfitting
- Hyperparameter γ plays a role in Radial Basis SVM Kernel. The higher the γ, the more the risk of overfitting
- One can search a grid of possible hyperparameters and then select the best hyperparameters based on Cross validation results
- SVR with Radial basis Kernel and Polynomial Kernel will take a long time to fit, if the dataset is large
- GCP's ML engine can be used to fit SVM on large datasets
- AWS Sage Maker can be used to fit SVM on large datasets
- TensorFlow could be used to distribute SVM fitting process across GPU's or TPU's