GUIDJIME ADINSI Ahouahounko

Machine learning

Supervised Learning

Supervised learning

Regression

- Linear Regression
- Regularised regression
- K-nearest Neighbors
- Decision Tree Regressor(CART)
- Support Vector Regression(SVR)
- AdaBoost
- Gradient Boosting method
- Random forest method
- Extras trees
- Artificial Neural Network (ANN)

Classification

- Logististic regression
- Linear discriminant analysis
- K-nearest neighbors
- Decision Tree Classifier(CART)
- Support Vector Classifier(SVC)
- AdaBoost
- Gradient boosting method
- Random forest method
- Extras trees
- Artifical Neutral network(ANN)

Pros and cons

Linear Regression(Ordinary Least Square)

Pros:

- Easy to understand and interpret;

Cons:

- Don't work well when there is nonlinear relationship between predicted and predictor variables;
- Prone to overfitting;
- May not handle irrelevant features well, when there is a large number of features ;
- Required the data to follow certain assumptions(absence of multicollinearity…)

Pros and Cons

Regularized regression

Pros:

- Prediction accuracy;
- Lasso regression not only helps in reducing overfitting , but also can in features selecting;
- Ridge regression shrinks the coefficient and helps to reduce the model complexity ;

Cons:

- Interpretation: A large numbers of predictors may complicate the interpretation or communication of big picture of results.

Pros and cons

Logistic regression

Pros:

- Easy to implement, good interpretability;
- Perfom very well on linearly separable classes;
- The model has small number of hyperparameters.

Cons:

- Overfit when provided with large members of features;
- Can only learn linear function;
- Less suitable to complex relationship;
- May not handle irrelevant features well, especially if the features are strongly correlated.

Pros and Cons

Support Vector Machine

Pros:

- Fairly robust against overfitting, especially in higher dimensional space;
- Handle the nonlinear relationships quite well, with many kernels to choose form.
- No distributional requirement for data

Cons:

- Inefficient to train and memory;
- Does not perfom will with large datasets ;
- Required the features scaling of the data;
- Having many hyperparameters and their meaning are often not intrutive;

Pros and Cons

K-Nearest Neighbors (KNN)

Pros:

- No training is involved and hence there is no learning phase;
- New data can be added seamlessly without impacting the accuracy of the algorithm;
- Intuitive and easy to understand;
- Handles multiclass classification and can learn complex decision boundaries;
- Robust to noisy data;
- No need to filter outliers.

Cons:

- Difficulty o choosen the distance metric;
- Perfom poorly on high dimensional datasets;
- Expensive and slow to predict new instances;
- Sensitive to noise in the to dataset ;
- Feature scaling(standardization, normalization ...) is required before

Pros and Cons

Linear Discriminant Analysis

Pros:

- Relatively simple model with fast implementation;

Cons:

- Required feature scaling and involves complex matrix operation;
- Some assumptions about the data : data are normally distributed, each attribute has the same variance and the values of each variable vary around the mean by the same amount on average;

Pros and Cons

Classification and Regression Trees(CART)

Pros:

- Easy to interpret and can adapt to learn complex relationships;
- Required little data preparation, data does not need to be scaled;
- Feature importance is built;
- Performed well on large datasets;
- Worked for both regression and classification problems…

Cons:

- Prone overfitting unless pruning is used;
- Can be very nonrobust: meaning that small changes in the training dataset can lead to quite major differences in the hypothesis function that gets learned;
- Has worse performance than ensemble models…

Machine learning

Unsupervised algorithms ....coming soon

...to be continued!!!