

**ELG5131 --- Graphical Models**

**Project 1**

**< Sum-Product Decoding >**

**Instructor:**

Professor Yongyi Mao

**Student Identification:**

Jianzhou Wang (Maxwell) --- 6696168

Rui Chen --- 6727535

Hui Zhao --- 7409618

**DATE:**

October 20th 2017

**Introduction**

Classifiers are important roles in machine learning, the goal of using a variety of classifiers is to identify different classes. A classifier is trying to be mining the data. The concept behind it is to set up a model of classification on the basis of data and train the date. On the other hand, it maps the data from the record to predict the next data set as well. First we will specify on classification and regression. Regression is to model the relationship between a continuous input variable x and a continuous target variable t. Classification will give us discrete form of target variables while the input variable may still be continuous. Logistic Regression, Recurrent (Recursive) Neural Networks, Convolutional Neural Networks, Support Vector Machine etc are all in the family of classifiers.

In this project, we are here applying to use three different classifiers (Logistic Regression, Support Vector Machine and K- Nearest Neighbors) to analyze and compare their efficiencies. Logistic regression will be primarily discussed in this project, this regression is based on the case of binary dependent variable, where the outcome is either 0 or 1. The important point here to note is that in linear regression, the expected value of the response variable are modeled based on combination of values taken by the predictors.

Logistic regression is able to

* Model the probabilities of a response variable as a function of some explanatory variables, e.g “success” of admission as a function of gender.
* Perform descriptive discriminate analyses such as describing the differences between individuals in separate as a function of explanatory variables
* Predict probabilities that individuals fall into two categories of the binary response as a function of some explanatory variables.
* Classify individual into two categories based on explanatory variables

Support Vector Machine (SVM) is a common check algorithm in machine learning. The principle is form linear to non-linear scenario. Meanwhile, it also constructs a hyper-plane or set of hyper-planes in a high or infinite-dimensional space, and a good separation is needed for such hyper-plane that has the largest distance to the nearest training-data point.

SVM is able to

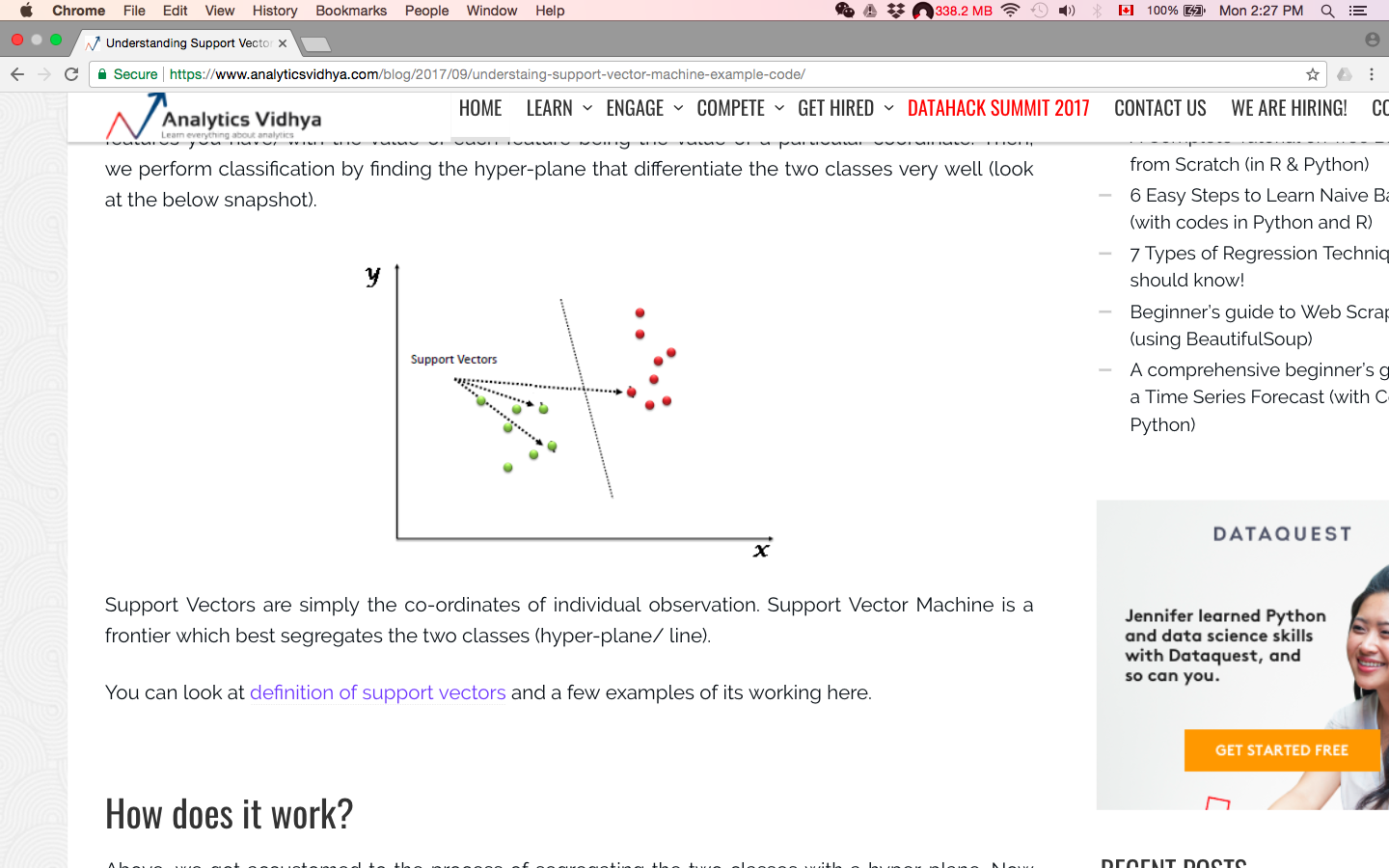
* Mapping the non-linear in low dimension to linear in high dimension. Therefore, the question can be resolved in linear in high dimension.
* Optimize the entire machine learning based on the minimized the experimental results to show that SVMs achieve higher accuracy than traditional query refinement schemes.

K-Nearest Neighbors (KNN) Algorithm is implemented by choosing the nearest data set and the most “look like” data which has number of k that belongs to the class. It is quite often to consider more than just one neighbors to the determine the class. Therefore, the k nearest neighbor are used in this case which refer to the KNN classification.

**Procedure**

*Support Vector Machine*

Support Vector Machine is also an outstanding supervised learning methods for classification, regression and outliers detection. In this algorithm, it plots each data item as a point in n-dimensional space with the value of each feature being the value of a particular coordinate and performs classification by finding the hyper-plane that differentiate the two classes.



Therefore, they are several benefits of support vector machines:

* Effective in high dimensional spaces.
* Often but not always effective when the number of dimensional is higher than the number of samples.
* Memory efficient by using a subset of training points in the decision function.
* A variety of different Kernel functions can be applied in this method .

The core technique of SVM is called kernel trick (function). As its mentioned before, it takes low dimensional input space and transform it to a higher dimensional space. It is mostly useful in non-linear separation problem. Thus it helps to resolve extremely complex data transformations, then fund out the process to separate the data based on the labels.

There are still a couple of disadvantages of support vector machines:

* If the number of features is much greater than the number of samples, avoid over-fitting in choosing Kernel functions and regularization term is crucial.
* SVMs do not directly provide probability estimates.

There are 4 different kernel functions we can use:

* Linear kernel SVM: <x, x’>
* Degree-d polynomial Kernel: ( <x, x’> + r)d. Where d is the *degree*, and r is *coef0*
* Radial basis function: exp( - ||x-x’||2). Where t is the keyword *gamma*, and t must be greater than 0
* Sigmoid kernel function: sigmoid (tanh( <x, x’> + r)), where t is specified by *coef0*

The function we choose is the third kernel function --- radial basis function. This is a popular function for approximation since SVM and other models employing the kernel trick do not scale well to large numbers of training samples or large numbers of features in the input space.

Let be a kernel function and its matrix could be define as . According to kernel functions’ properties, there exists a map which satisfies and the classification vector will be rewritten in this form [6]

So the optimization problem could be rewritten in this form

Substitute the radial basis function we mentioned before

is dependent on the users’ inputs, but it must be greater than 0