Interactive Machine Learning Interface

Lilong Jiang(jiang.573)

1 Abstract

2 Introduction & Motivation

3 Related Work

[1] shows an interactive interface for machine learning. We use a classification Pima Indians Diabetes Data Set from UCI [7].

4 Work

The overall space is divided into two parts: the left part shows what kind of machine learning method to run and parameters associated with the algorithm. Also users can choose different dataset and it will show the attributes of current manipulated dataset. In general, we implement three machine learning algorithms in the frontend using javascript: K-means (Figure 4), linear regression (Figure 4) and support vector machine (SVM) (Figure 5). Table 4 summarizes operations that users can perform with our interface.

| algorithms	operations
k-means	(1) Select initial centers or generate centers randomly (2) Adjust number of clusters on-line (3) Colors shows different clusters
linear regression	(1) Adjust function's intercept by moving the line (2) Adjust function's slope and intercept by dragging end points (3) Dynamically show the function and square loss
svm	(1) Select kernel (2) Select which features to use by dragging into or out space (3) Dynamically show accuracy, precision, recall and confustion matrix

Table 1: Operations
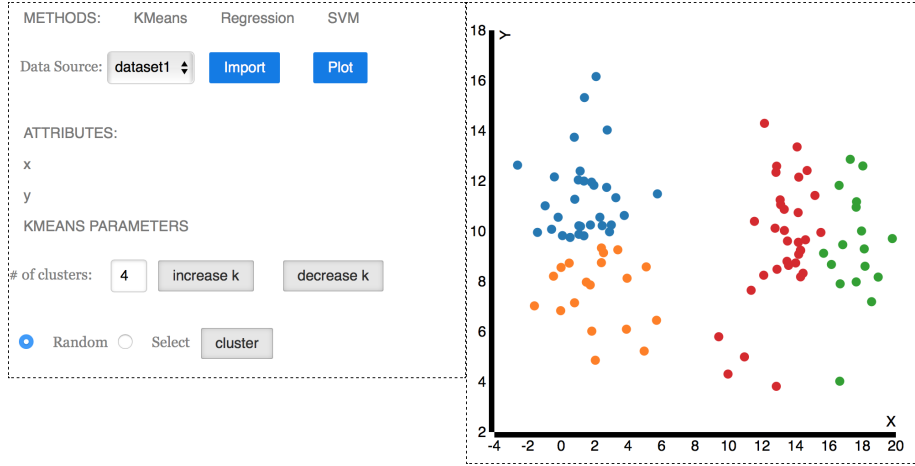
Figure 1: K-means

5 Performance Experiments

In this section, we focus on performance experiments since in the interactive environment, it is necessary to keep an interactive performance.

Computer Configuration:

Dataset: For SVM, we use Skin Segmentation Data Set from UCI [6]. There are totally 3 numerical attributes and 2 classes. We randomly sample from both classes.

Results:

6 Limitations & Conclusions

In this project, we focus on usability point. There are several limitations for this project. The first is interactive performance. The second is project issue for high-dimensional dataset. *Projection* is a technique that maps high-dimensional dataset into a small set of dimensions. Some works use dimension reduction, that create latent dimensions that summarize dataset. For example, Principal Component Analysis (PCA) [4], Multidimensional Scaling (MDS) [5]. This method can reveal hidden variables. However, the generated dimensions are usually less intuitive to users. Another method is called feature selection, which selects subset of features to explore. Interactive feature selection is often used to identify relationship between features [3, 8] and aid user remove redundant features and choose appropriate features. Recently, there is a work trying to allow users to craft their own projection function [2]. All these methods can be easily incorporated into our project.
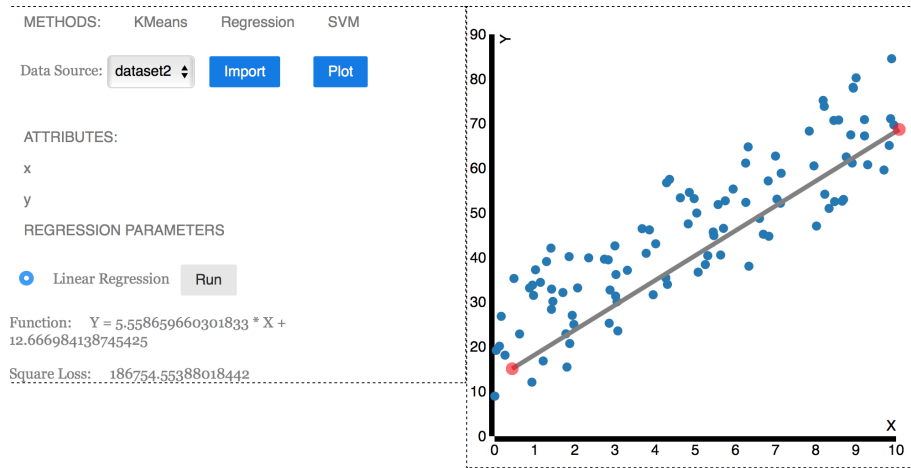
Figure 2: Linear Regression

References

[1] Andrew Crotty, Alex Galakatos, Emanuel Zgraggen, Carsten Binnig, and Tim Kraska. Vizdom: interactive analytics through pen and touch. *Proceedings of the VLDB Endowment*, 8(12):2024–2027, 2015.

[2] Michael Gleicher. Explainers: Expert explorations with crafted projections. *TVCG*, 2013.

[3] Diansheng Guo. Coordinating computational and visual approaches for interactive feature selection and multivariate clustering. *Information Visualization*, 2003.

[4] Ian Jolliffe. *Principal component analysis*. 2002.

[5] A Mead. Review of the development of multidimensional scaling methods. *The Statistician*, 1992.

[6] Abhinav Dhall Rajen Bhatt. Skin segmentation dataset. *UCI Machine Learning Repository*.

[7] Jack W Smith, JE Everhart, WC Dickson, WC Knowler, and RS Johannes. Using the adap learning algorithm to forecast the onset of diabetes mellitus. In *Proceedings of the Annual Symposium on Computer Application in Medical Care*, page 261. American Medical Informatics Association, 1988.
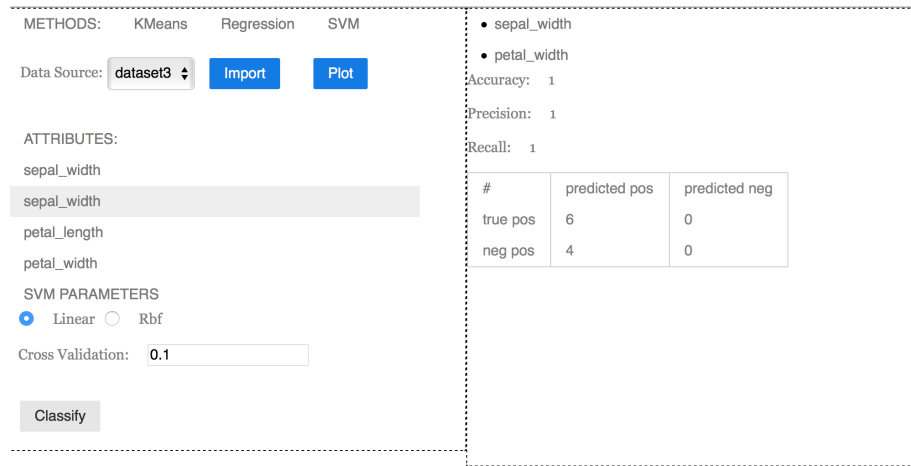
Figure 3: SVM

[8] Jing Yang et al. Value and relation display for interactive exploration of high dimensional datasets. In *INFOVIS*, 2004.

METHODS: KMeans   Regression   SVM

Data Source: dataset3 ⬍   Import   Plot

ATTRIBUTES:
sepal_width
sepal_width
petal_length
petal_width
SVM PARAMETERS
● Linear ○ Rbf
Cross Validation: 0.1

Classify

• sepal_width
• petal_width
Accuracy: 1
Precision: 1
Recall: 1

#	predicted pos	predicted neg
true pos	6	0
neg pos	4	0

Figure 4: SVM

5