SVM拥有高准确率，为避免过拟合提供了很好的理论保证，而且就算数据在原特征空间线性不可分，只要给个合适的核函数，它就能运行得很好。其优点在于可以解决高维问题，即大型特征空间，能够处理非线性特征的相互作用，无需依赖整个数据，可以提高泛化能力。而同时，其也存在很多缺点，如当观测样本很多时，效率并不是很高，对非线性问题没有通用解决方案，有时候很难找到一个合适的核函数，对缺失数据敏感。

libsvm中自带了四种核函数：线性核、多项式核、RBF以及sigmoid核，如果样本数量小于特征数，那么就没必要选择非线性核，简单的使用线性核就可以，如果样本数量大于特征数目，这时可以使用非线性核，将样本映射到更高维度，一般可以得到更好的结果，如果样本数目和特征数目相等，该情况可以使用非线性核，原理和第二种一样。

在跑本模型的分类运用时，无论是在对数据进行预处理和特征选择与否的情况下，我们发现在运用RBF核的情况下准确率会有所提高，但是并不是很理想，准确率较低。

而在跑本模型的回归运用时，我们发现在运用sigmoid核的情况下，准确率较高，结果比较理想。

SVM has a high accuracy rate, which provides a good theoretical guarantee to avoid overfitting, and even if the data is linearly inseparable in the original feature space, if a suitable kernel function is given, it can run well. Its advantage is that it can solve high-dimensional problems, that is, large feature spaces, and can handle the interaction of nonlinear features without relying on the entire data, which can improve the generalization ability. At the same time, it also has many shortcomings, such as when there are many observation samples, the efficiency is not very high, there is no general solution to nonlinear problems, sometimes it is difficult to find a suitable kernel function, and it is sensitive to missing data.

Libsvm comes with four kernel functions: linear kernel, polynomial kernel, RBF, and sigmoid kernel. If the number of samples is less than the number of features, then there is no need to choose a nonlinear kernel. Simply use a linear kernel. If the number of samples is greater than the number of features Number, at this time, you can use a nonlinear kernel to map the samples to a higher dimension, and generally get better results. If the number of samples is equal to the number of features, you can use a nonlinear kernel in this case. The principle is the same as the second one.

When running the classification of this model, whether it is in the case of data preprocessing and feature selection or not, we found that the accuracy rate will be improved when the RBF kernel is used, but it is not very ideal. The accuracy rate lower.

When running the regression application of this model, we found that in the case of using the sigmoid kernel, the accuracy rate is higher, and the result is more ideal.

Description of SVM:

SVM是一类按监督学习方式对数据进行二元分类的广义线性分类器，其决策边界是对学习样本求解的最大边距超平面，可以将问题化为一个求解凸二次规划的问题。与逻辑回归和神经网络相比，支持向量机，在学习复杂的非线性方程时提供了一种更为清晰，更加强大的方式。

具体来说就是在线性可分时，在原空间寻找两类样本的最优分类超平面。在线性不可分时，加入松弛变量并通过使用非线性映射将低维度输入空间的样本映射到高维度空间使其变为线性可分，这样就可以在该特征空间中寻找最优分类超平面。

SVM is a kind of generalized linear classifier for binary classification of data according to the supervised learning method. Its decision boundary is the maximum margin hyperplane for solving the learning samples, which can turn the problem into a problem of solving convex quadratic programming. Compared with logistic regression and neural networks, support vector machines provide a clearer and more powerful way to learn complex nonlinear equations.

Specifically, when linearly separable, find the optimal classification hyperplane for two types of samples in the original space. When linearly inseparable, add slack variables and use nonlinear mapping to map the samples of the low-dimensional input space to the high-dimensional space to make it linearly separable, so that the optimal classification hyperplane can be found in the feature space.

Parameter of SVR：

SVM回归算法称为SVR，我们从多个实验数据中发现，SVR模型中的kernel，gamma， coef0， C核函数会对实验结果产生比较明显的影响。经过多次实验和MATLAB数据分析，kernel='sigmoid', gamma='auto', coef0=5.0, C=2.0时将出现最优值，因此取本次最优值作为参数训练Random Forest模型。

The SVM regression algorithm is called SVR. We found from multiple experimental data that the kernel, gamma, coef0, and C kernel functions in the SVR model will have a more obvious impact on the experimental results. After many experiments and MATLAB data analysis, the optimal value will appear when kernel='sigmoid', gamma='auto', coef0=5.0, C=2.0, so this optimal value is used as the parameter to train the Random Forest model.

Results：

The MAE of SVM is 20.66

Explanation and presentation of the results obtained:

我们综合使用了各种数据预处理，特征选择的方法（对比结果见表：） 经过对比，我们最终采用了selectFromModel特征选择的方法，使用SVM回归模型，得到了这样的结果。

We comprehensively used various data preprocessing and feature selection methods (see table for comparison results:) After comparison, we finally adopted the selectFromModel feature selection method and used the SVM regression model to obtain such results.