ECEN765 Machine Learning with Networks -- Course Project Report

Name: Liuyi Jin 225009797

Class: Application

Abstract: Type wells, which are typically well production profiles based on analysis of existing well histories, are growing in acceptance in the industry as a means of forecasting production in low-permeability reservoirs. Although the conceptual understanding is becoming clearer, there are still many challenges when we try to apply type wells in practice, chief among them being how to apply the type wells that accurately reflect original well production data to the field application. At the same time, we will adopt two machine learning principles, Neural Networks (NNet) and Support Vector Machine (SVM), as a significant method to classify type wells. By using production data from the Barnett Shale, this project will be a trial to apply both type well construction strategies and machine learning principles, to traditional oil and gas production data analysis. It also gives an indication that can be further used in financial investment decision making process in oil and gas industry.

1. Introduction

In the petroleum engineering industry, some researchers are working on how to accurately predict the production of newly drilled wells according to the analysis of earlier production from older wells. In the past, when people try to predict production of conventional oil and gas reservoirs, they adopted deterministic prediction models. Those methods are derived from a fundamental equation in fluid mechanics area called Darcy's Law. However, since 2008, unconventional reservoirs, which cannot be accurately predicted directly by using previous deterministic methods, are becoming important in the world energy market. Researchers then began to find innovative ways to accurately predict this kind of new unconventional reservoirs. Soon people in this area found that prediction methods with statistics principles underlined are very suitable in this particular occasion. Type well method is one of those methods.

Type wells, which are typical well production profiles based on analysis of existing well histories, are growing in acceptance in the industry as a mean of forecasting production in low-permeability reservoirs. The core idea of constructing type wells is to construct a well representing a set of wells that are being recognized as "analogous" in a particular geologic area. It aims to extract the inherent characteristics of multiple wells in a certain geologic area by producing a family of constructed wells which should be representative enough to represent all wells in an area of interest. The approach commonly employed in industry to construct type wells is to arithmetically average the production histories of a set of producing wells in the field. I will also take this method in this particular project.

Artificial Neural Networks, which are currently gaining much acceptance in the world. More and more people tends to employ this method within their own researches. There are several mathematical algorithms to train the neural networks. For example, steepest gradient descent, conjugate gradient descent, stochastic gradient descent (with momentum) are included in popular algorithms to train the target neural networks.

Support Vector Machines, SVM, is the classical machine learning algorithm. SVM deals with the problems that requires soft margin. It can be used in both the linear problems and problems with higher non-linearity. The underlying principle of SVM is the sequential minimal optimization which make sure to find the optimal solution globally.

2. Methodology

Our project is now aiming to train the given data set with given type wells, that is to say, we are going to classify the wells into several types using NNet and SVM.

2.1 Dataset

In the petroleum production field, production data are recorded as time series data, rate versus month precisely speaking. Those data are given on the website https://info.drillinginfo.com, which is an official website specifically focus on providing nation-wide oil and gas production data. I picked up 200 wells that are active in production.

The 200 wells have different starting production time in month, but share the same ending production time in May 2017, which results in the different feature length for them. Our first step is to label those 200 samples. For each well in the field, there is a typical production profile as shown in the figure 1. The green dots showed the trend of declining rate of production, the red dot lines gives the typical trend for a cumulative production. With the time increasing, the production would decrease below a threshold, which is called abandon rate. Normally speaking, the life of one particular well should last for over 30 years before it decreased to its abandon rate. This gives us a clue to label the 200 wells.

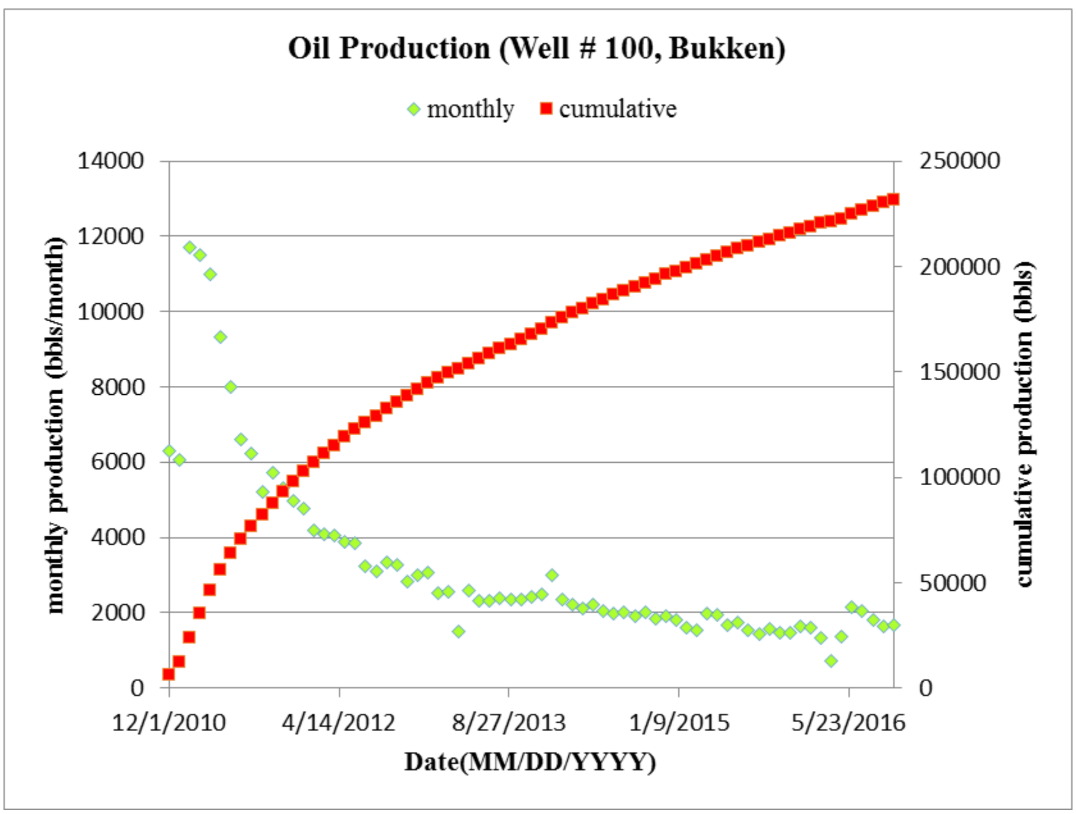

Figure 1 Typical Well Production Profile

2.2 Forecasting

As we see from the figure 1, we have only limited data points (i.e. less than 360 months). For the sake of reliably labelling the available dataset, we used some physical models in petroleum industry to make the forecasting, the forecasting data would be further used in the labelling stage. Given that the data are all produced in the unconventional reservoir, I expected its flow pattern changing from linear transient flow to boundary dominant flow. With this taken into consideration, I tried several prediction models that are popular in the petroleum industry.

<u>Power Law</u> (Ilk, Rushing et al. 2008)

$$D = D_\infty + D_1 t^{-(1-n)}$$

$$q = \hat{q}_i exp\left(-D_\infty t - \widehat{D}_i t^n\right)$$

Where:

D_1 is the decline constant "intercept" at 1 time unit, 1/days

D_∞ is the decline constant at infinite time, 1/days

n is the time exponent, unitless

$\hat{q}_i$ is the rate intercept bbl/day or Mcf/day

<u>Stretched Exponential</u> (Valko and Lee 2010)

$$q = q_0 exp\left[-\left(\frac{t}{\tau}\right)^n\right]$$

Where:

q is the production rate at any time

q_0 production rate at time = 0

τ is the characteristic time parameter

n is the time exponent, unitless

<u>Duong's method</u> (Duong, 2011)

$$\frac{q}{G_p} = at^{-m}$$

Where:

G_p is the cumulative production

q is the production rate, vol/day

t is the time, days

a & m are constants

Using the software, ValNav, I can easily choose the best fit from the three models for each well to do the forecasting. Figure 2 and figure 3 gave the GUI interface for the ValNav and BestFit, respectively.

Since ValNav has wrapped up the fitting models for us, so we directly input the raw data, which is extracted from the drillingInfo website. Sometimes, the ValNav may give us data more than 360 months, but we only use the 360 months data. Those wells are given in the log_prob_plot.xls file.

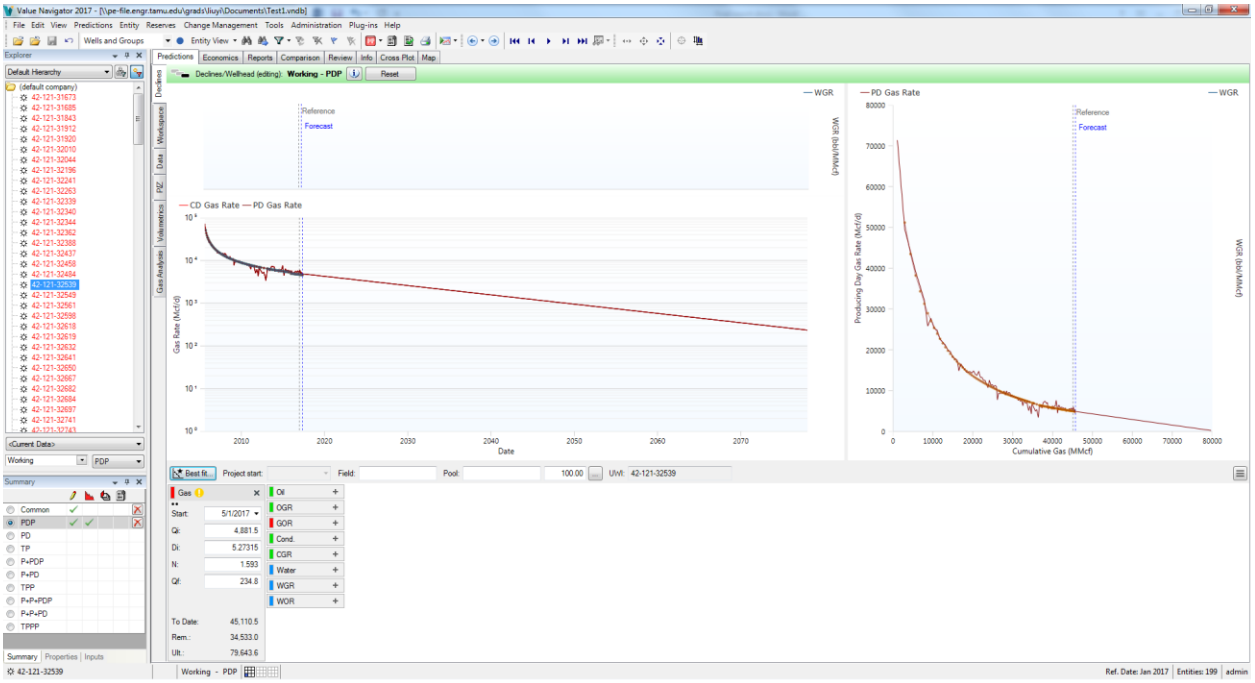

Figure 2 ValNav Interface

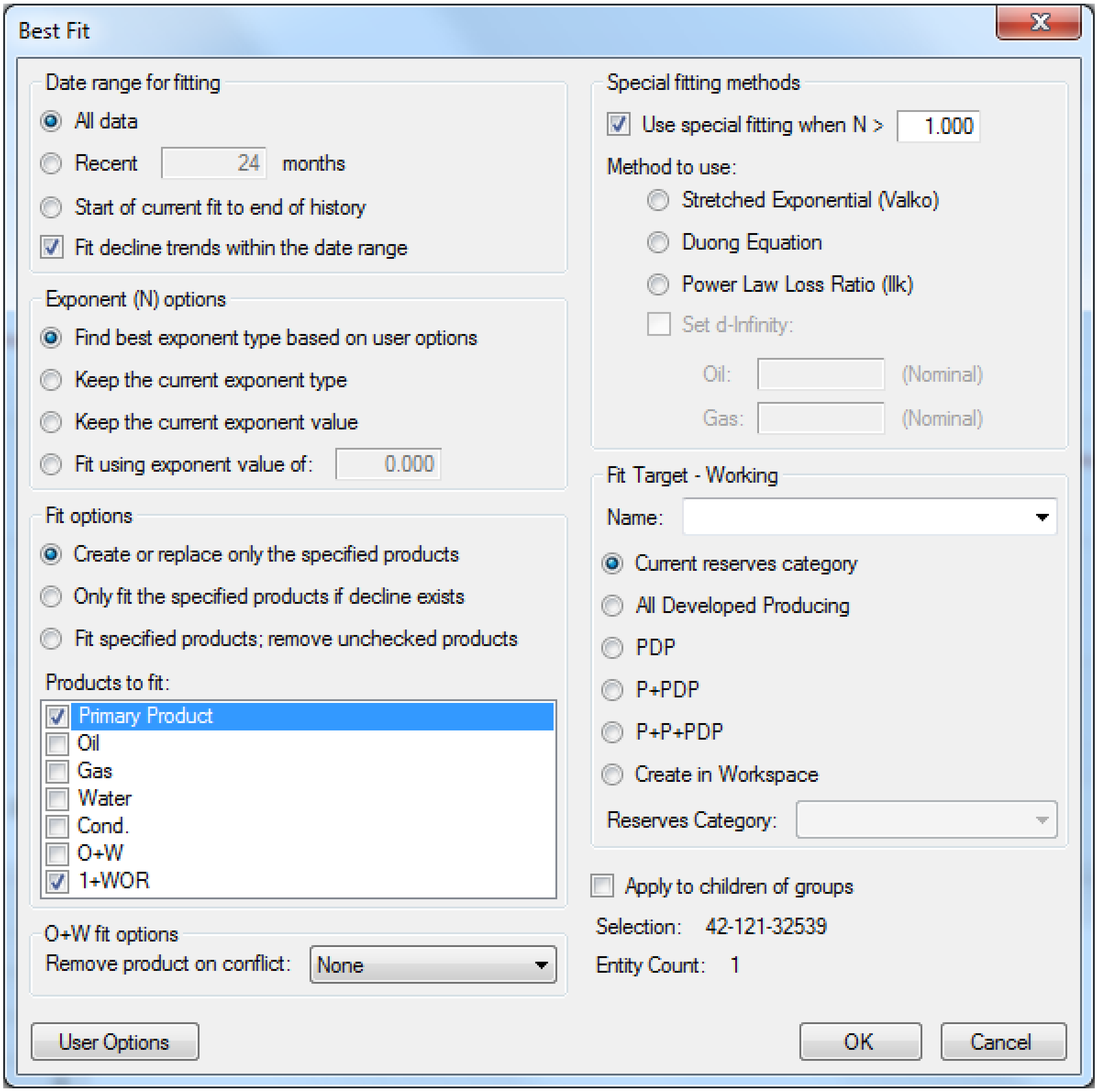

Figure 3 Best Fit interface

2.3 Labelling

With the aid from ValNav, we can get the estimate ultimate recovery for each well and we extract the 360 months data from the first production month, i.e. the feature space for 200 well samples becomes 360. In type wells concepts, the labels given to each well sample are called types. The types are classified according to their corresponding estimate ultimate recovery which is the last value on the red dotted line in figure 1. With 200 estimate ultimate recovery values given, we sort them accordingly. In this case, each estimate recovery value for each well is corresponding to a “less than probability”. Since the estimate ultimate recovery is in log normal distribution, so I plotted the EUR distribution in a lognormal paper as shown in figure 4.

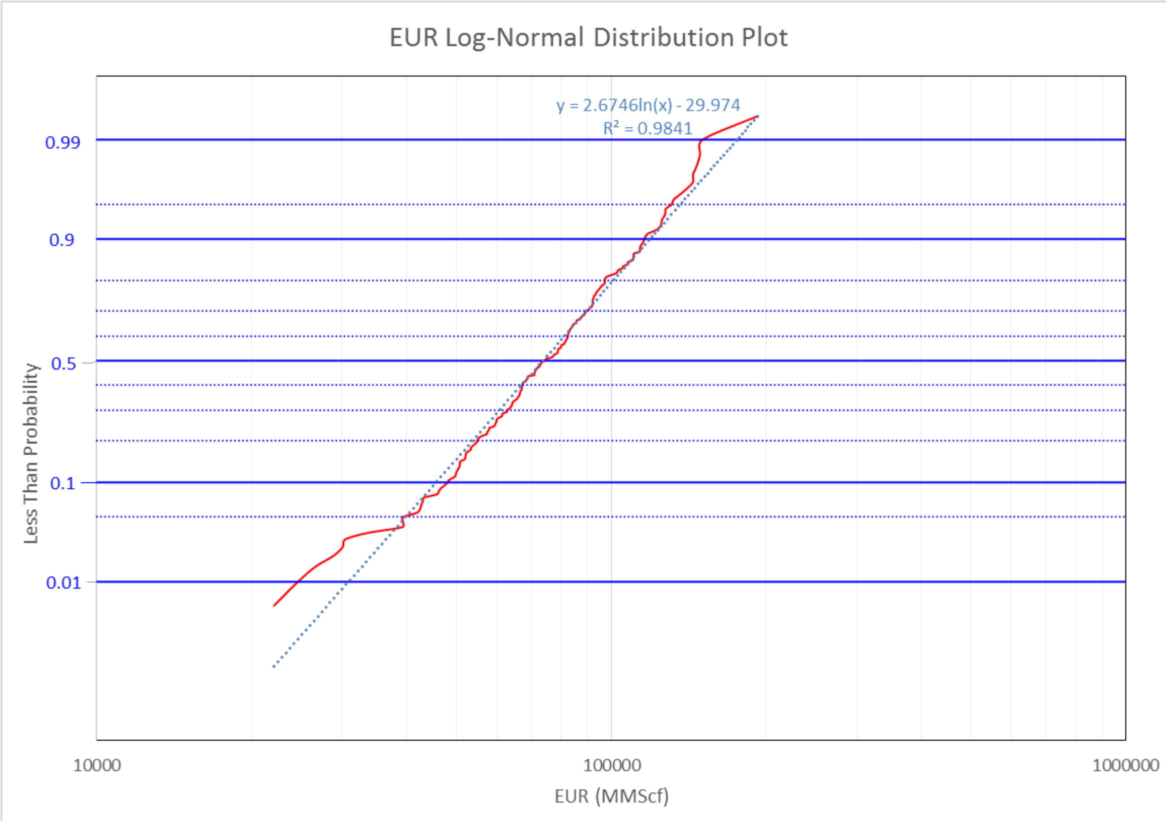

Figure 4 EUR LogNormal Distribution

Problem comes to how to divide the samples into different types. In this project, I separated the well samples into 4 types. This separation schema is commonly employed in petroleum literatures. Ba careful that P10, P50 and P90 values are referring to greater than values, which is intuitively contrary to less than probabilities shown in the plot in figure 4.

Type1 - Above P90, the P90 corresponding EUR is 48062.86016 in figure 4, probability 0.1

Type 2 – P50 – P90, the P50 corresponding EUR is 74265.31784 in figure 4, probability 0.5

Type 3 – P10 – P50, the P10 corresponding EUR is 115750.2296 in figure 4, probability 0.9

Type 4 – Below P10, the largest EUR is 192576.9152 in figure 4, highest point in the graph

After the type wells in a particular geologic area are constructed, they are claimed to be representative enough with a certain confidence. We train the rest data to fit them into one of those type wells. We will choose artificial neural network and support vector machine are our primary training methods to classify newly drilled wells with limited production history into one of our types.

2.4 Data Training

Training with neural network

Given a large set of data (in our case, they are 200 production wells production) samples. We already know that they have been labelled with corresponding class label (type 1, type 2, type 3, and type 4). The number of types are finite. Our goal is to train a neural network with a subset of the original 200 data samples. Here, I need to clarify that the input data into the input layer is actually the raw data from drillingInfo website rather than the forecasted data with 360 features for all data samples, the forecasted data is only used when I tried to give the labels to samples. To fit those data into the constant number of neurons in input layer, I extrapolate some data samples to give the same number of features.

In addition, given the fact that I have only 200 samples, each with constant number of features. I chose a subset of the 200 samples as the training set with the rest of the data as the test samples (actually I used 100 sample data to train, and used the left 100 as the test set). The expected result is a trained network with its weights updated, which can be used to correctly classify new coming wells with limited time of history.

The point of the classification is that: as I classify one test sample into type 1, I could claim that I have 90 percent confidence to have estimate ultimate recovery from that specific well larger than 48062.86016 MMScf; as I classify one test sample into type 4, I could claim that I have only 10 percent confidence to have estimate ultimate recovery from that specific new coming well larger than 192576.9152 MMScf. The gives the importance of the type well in reserve evaluation and financial decision making for oil and gas companies.

As for the algorithm used in this project, it would be stochastic gradient descent. The process can be graphically represented as shown in figure 5. The figure below did not show the back propagation.

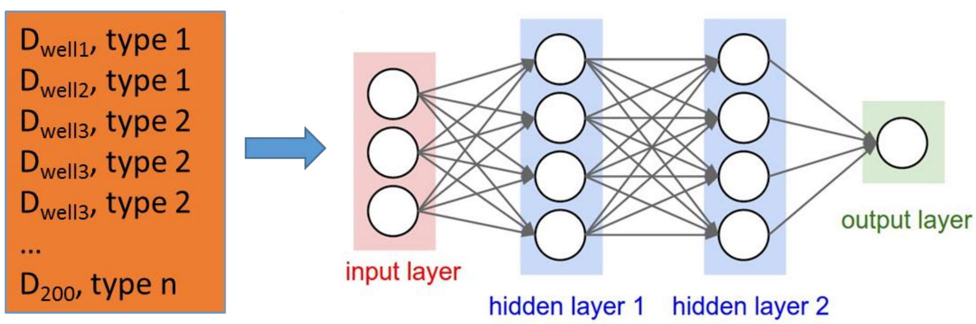

Figure 5 Schematic Representation of Type Well Data Training (http://cs231n.github.io/neural-networks-1/)

This project will be a trial to apply both type well construction strategies and machine learning principles, to our traditional oil and gas production data analysis. We will further apply probabilistic theory to classify the undrilled wells.

Training with support vector machines

The data preprocessing is mainly the same with neural networks elaborated above. I have 200 samples with each have constant feature (feature dimension is not 360). With the aid from the sklearn package in python, we could easily set the C values to have different soft margin SVM.

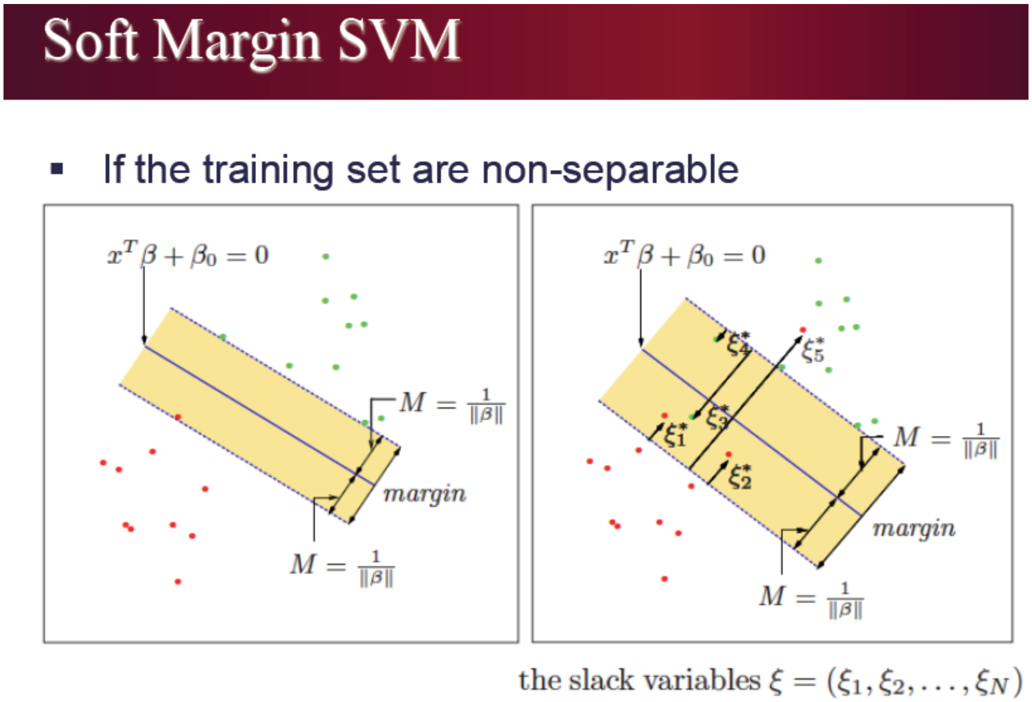

Figure 6 Soft and hard margin SVM (ECEN765 Machine Learning with Networks, Xiaoning Qian, Fall 2017)

3. Results and Conclusions

As shown in figure 7, I trained the data and had the results. For the 100 training data, I got relatively more accurate accuracy results than test cases. In terms of the training algorithms, SVM performance is far better (C = 0.2, linear kernel). This should be due to the fact that SVM adopts sequential minimization as its underlying optimization principle, which helps the SVM find the globally optimal solution. We could infer from the results shown in figure 7 that SVM is more suitable in terms this specific classifying problem.

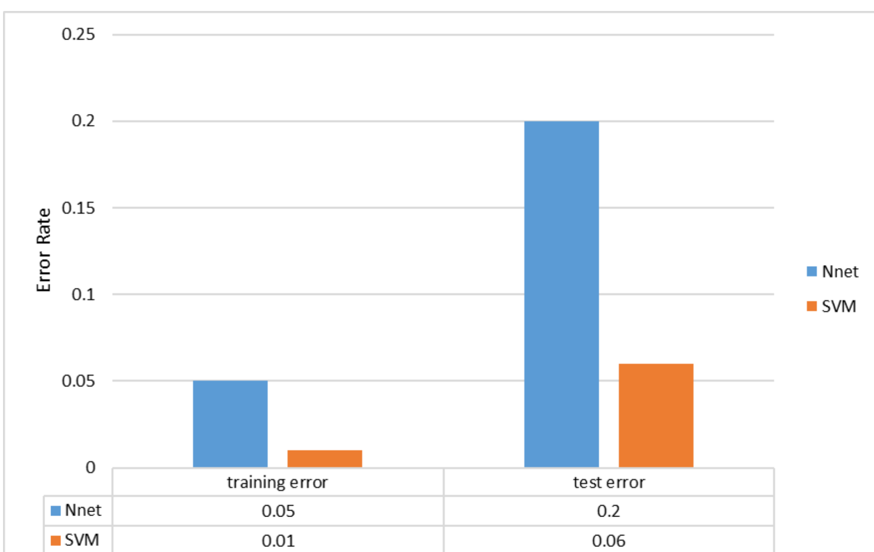

Figure 7 Data Training Result

4. References

D. IIK, J.A. Rushing, A.D. Perego, T.A. Blasingame, Exponential vs. Hyperbolic decline in tight gas sands: understanding the origin and implications for reserve estimate using Arp's decline curves, Presented at the SPE Annual Technical Conference and Exhibition, 21-24 September (2008, January 1), 10.2118/116731-MS, Denver, Colorado, USA

Valko, P. P., & Lee, W. J. (2010, January 1). A Better Way To Forecast Production From Unconventional Gas Wells. Society of Petroleum Engineers. doi:10.2118/134231-MS

A.N. Duong, 2011, Rate-decline analysis for fracture-dominated shale reservoirs, SPE Reserves Evaluation Engineering, 14(03) (2011), pp.377-387, 10.2118/137748-PA

http://cs231n.github.io/neural-networks-1/

ECEN765 Machine Learning with Networks, Xiaoning Qian, Fall 2017