Continuous blood pressure measurement system without the blood pressure cuff

LAI Jianfa, CHEN Xiao
Hong Kong Baptist University

April 24, 2019

Abstract

Continuous blood presure measurement system is significantly demanded by the patients with hypertension and accompanying heart problems, especially during the midnight. And the standard blood pressure measurement needs the professional nurses and the blood pressure cuff which can not continuously bound on the arm, influencing seriously on life quality of patients. This paper talks about serveral methods (models) to solve these problems by using PhotoPlethysmoGraphy (PPG) data. We will do the classification problems or the prediction with the PPG data. Our aim is to classify Hypertensives in Hypertensive group with over 85% accuracy and 75% accruracy for normal people in normal group. Or we will predict the blood presure within 5mmHg error and the standard deviation is less than 8mmHg.

1 Introduction

1.1 Background

1.1.1 PPG data

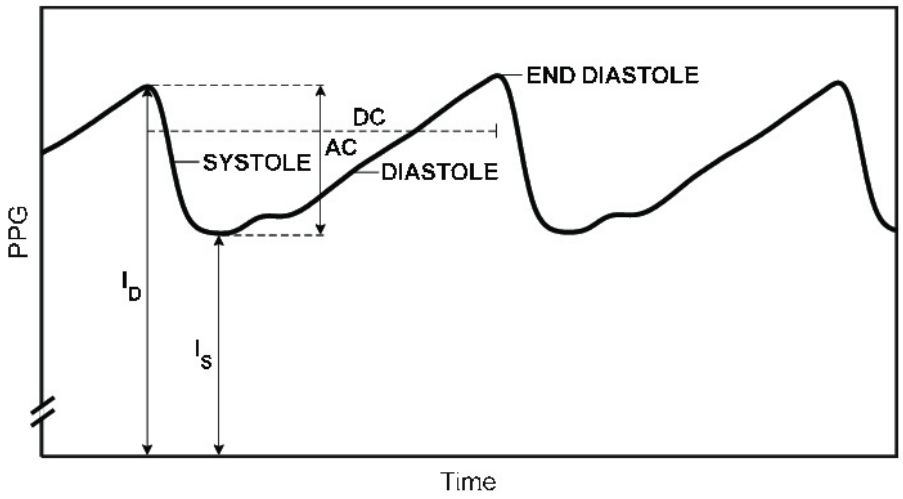

Figure 1: PPG data

The PPG waveform comprises a pulsatile ('AC') physiological waveform attributed to cardiac synchronous changes in the blood volume with each heart beat, and is superimposed on a slowly varying ('DC') baseline with various lower frequency components attributed to respiration, sympathetic nervous system activity and thermoregulation. (Allen J., 2007) PPG can be simply considered as a function of $t, f(t)$. There are several features or signals, like AC,DC, the maximum(ID) and the minimun(IS), the maximum of the $f(t)$,etc. All these features can also generate other features. For example, one feature divided by another feature, or one multiplies another, or we can simply take log for each feature. This step is significant. We can use simple models with the new feature set to approach the true model, even if it is non-linear.

1.1.2 Problem Identification

The new feature set is considered as independent variables x, and p the dimension of x. For the classification, $y = \{0, 1\}$. For $y = 0$, it means the sample is the normal person. For y=1, the sample is the Hypertensive.

Our aim is to classify Hypertensives in Hypertensive group with over 95% accuracy and 75% accruracy for normal people in normal group. For the prediction, y is the 2 dimension vector, the systolic blood pressure(sbp) and diastolic blood pressure(dbp). The result of the prediction should be within 5mmHg error and the standard deviation is less than 8mmHg.

1.2 Method Introduction

1.2.1 Linear Regression

Linear Regression is a widely used approach to predicting a quantitative response with some related variables. Linear regression can reveal the linear relationship between the response and one or more predictors. It assumes that the response y linearly depends on p distinct predictors . Mathematically, this linear regression model can be written as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \epsilon \tag{1}$$

where x_i denotes the ith predictor,β_i quantifies the association between the response y and that predicor. and ϵ denotes a random error with zero mean. Given the estimator $\hat{\beta}_i$, the prediciton for y will be

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_p x_p \tag{2}$$

Let $\{(x_{i1}, x_{ip}, y_i) : i = 1, 2 \ldots, n\}$ represent n observation pairs. The sum of squared residuals can be written as

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{3}$$

$$= \sum_{i=1}^{n} (y_i - \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \ldots + \hat{\beta}_p x_{ip})^2 \tag{4}$$

The smaller the squared error, the more accurate the model fitting is. Thus, we choose $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$ that minimize RSS.

1.2.2 Expectation Maximization Algorithm

Sometimes a dataset includes missing values or assuming the existence of further unobserved variables can make the model formulated more simply. In these cases, The EM algorithm can be used for finding maximum likelihood parameters of statistical models. Let $\boldsymbol{X}$ represents the observed data and $\boldsymbol{Z}$ represents the unobserved data. $\boldsymbol{\theta}$ is a vector of unknown parameters.

The maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$ can be derived by maximizing the marginal likelihood of $\boldsymbol{X}$

$$L(\boldsymbol{\theta}; \boldsymbol{X}) = p(\boldsymbol{X}|\boldsymbol{\theta}) = \int p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\theta}) d\boldsymbol{Z} \tag{5}$$

However, integration is often not intractable or direct maximization of $L(\boldsymbol{\theta}; \boldsymbol{X})$ is often difficult. If we observed Z, the likelihood function would be $L(\boldsymbol{\theta}; \boldsymbol{X}, \boldsymbol{Z}) = p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\theta})$. Since we can't observe Z, we can calculate the expected value of $\log L(\boldsymbol{\theta}; \boldsymbol{X}, \boldsymbol{Z})$ with respect to the current conditional distribution $p(\boldsymbol{X}|\boldsymbol{Z}, \boldsymbol{\theta}^{(t)})$ first. This step is called E step.

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = E_{p(\boldsymbol{X}|\boldsymbol{Z}, \boldsymbol{\theta}^{(t)})}[\log L(\boldsymbol{\theta}; \boldsymbol{X}, \boldsymbol{Z})] \tag{6}$$

Then we can find the parameters by maximizing this quantity:

$$\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) \tag{7}$$

This is called M step.

Expectationmaximization Algorithm is an iterative method. We first initialize the parameters $\boldsymbol{\theta}$ to some random values. And then we repeat E step and M step until convergence.

Expectation-maximization tries to improve $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ instead of directly improving $\log L(\boldsymbol{\theta}; \boldsymbol{X})$.The following deductions show the feasibility of doing so.

We can write

$$\log p(\boldsymbol{X}|\boldsymbol{\theta}) = E_{p(\boldsymbol{X}|\boldsymbol{Z}, \boldsymbol{\theta}^{(t)})}[\log p(\boldsymbol{X}, \boldsymbol{Z})|\boldsymbol{\theta}] - E_{p(\boldsymbol{X}|\boldsymbol{Z}, \boldsymbol{\theta}^{(t)})}[\log p(\boldsymbol{X}|\boldsymbol{Z}), \boldsymbol{\theta}] \tag{8}$$

The gain for each iteration is

$$\log p(\boldsymbol{X}|\boldsymbol{\theta}) - \log p(\boldsymbol{X}|\boldsymbol{\theta}^{(t)}) \tag{9}$$

$$= Q(\boldsymbol{\theta}|\boldsymbol{\theta}) - Q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t)}) + E_{p(\boldsymbol{X}|\boldsymbol{Z}, \boldsymbol{\theta}^{(t)})} \frac{[\log p(\boldsymbol{X}|\boldsymbol{Z}), \boldsymbol{\theta}^{(t)}]}{[\log p(\boldsymbol{X}|\boldsymbol{Z}), \boldsymbol{\theta}]} \tag{10}$$

$$\geqslant Q(\boldsymbol{\theta}|\boldsymbol{\theta}) - Q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t)}) \tag{11}$$

$$\geqslant 0 \tag{12}$$

That is to say, choosing $\boldsymbol{\theta}$ to improve $Q(\boldsymbol{\theta}|\boldsymbol{\theta})$ could make the marginal likelihood $p(\boldsymbol{X}|\boldsymbol{\theta})$ non-decreasing.

1.2.3 Neural Network

Neural networks are the kernel technique for deep learning. A neural network can be viewed as a model consisting of a network structure composed of neurons.

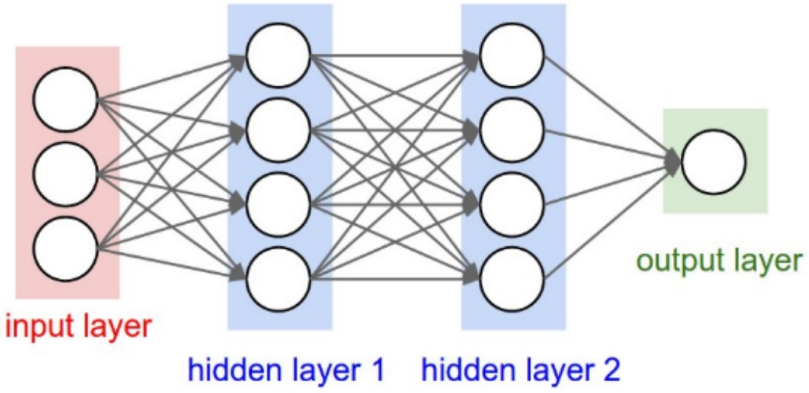

Figure 2: A neural network with 2 hidden layers

There are 3 types of layers in neural networks: input layer, hidden layer, and output layer. Neurons are connected by a weight. All inputs are summed according to this weight. This activity is called a linear combination. Finally, an activation function is used to get the outputs. If there are hidden layers, the outputs could be units in the hidden layers. Suppose there is one hidden layer. For input variables $x_1, x_2 \ldots, x_p$, their M linear combinations are

$$a_j = \sum_{i=1}^{p} w_{ji} x_i + w_{j0} \tag{13}$$

Then we can get M output units $z_j = h(a_j)$ though activation function $h(\cdot)$. These units are called hidden units. Then we can view these hidden units as input variables and repeat the process mentioned above. Finally,we can get the output $y_k, k = 1, 2, \ldots, K$. Given the loss function, we can optimize the parameters in the neural network. (Nasser M. Nasrabadi., 2007)

Neural networks are used in various fields, such as regression analysis, classification, and other applications where they can be trained via a dataset.

2 Methodology

2.1 Linear Logistic Regression

Linear Regression is easy and powerful method to test the data and get the sense of the understanding of the data. Since our question is classification problems, we decide to use the Linear Logistic Regression. Mathematically, the Linear Logistic Regression model can be written as

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \epsilon \tag{14}$$

$$y = \frac{1}{1 + \exp(-z)} \tag{15}$$

if $y > t$, then it is hypertensive. Otherwise, it is normal. Since the data set is unbalance(Less than 15% samples of the data are Hypertensive samples), the threshold is not longer 0.5. Here we let t be the threshold. We can control the t to adjust the result.

This method, of course, is easiest way to get the preliminary result and it can be seen as the base line.

2.2 Variational bayesian Expectation Maximization

The reason why we use the EM algorithm is that we believe that there must be something controlling the blood pressure that we can not detect, like the effects on the gene level or the effects on the organ level. Since all these random effects are impossible to figure out, we just make all the people as a latent variable, i.e, if the sample comes from a specific person, the latent variables of the sample are 1 in this person label, and 0 in other people label. So the latent variables of the data, Z, is the 0-1 vector, with over 3000 dimensions(there are over 3000 peoples). Thus due to this believe, we do need to use EM algorithm.

When we use the Expectation Maximization Algorithm, we find that it is difficult to calculate the prior distribution. However, if we make the assumption that each latent variables are independent, it will be easier to calculate the prior distribution. The algorithm is in the following.

$$\log \frac{\pi_{ij}}{1 - \pi_{ij}} = \eta_{ij} = x_{ij}^T \boldsymbol{\beta} + z_{ij} u_i, \tag{16}$$

where $z_{ij} = 1$, random effects $u_i \sim \mathcal{N}(0, \sigma^2)$, N individuals,each individuals have J_i runs.

$$Y_{ij} \sim Bernoulli(\pi_{ij}) \tag{17}$$

$$P(Y_{ij}|\boldsymbol{\beta}, u_i) = \exp(Y_{ij}\eta_{ij})\sigma(-\eta_{ij}) \geq \exp(Y_{ij}\eta_{ij})\sigma(\epsilon_{ij})\exp[-\lambda(\epsilon_{ij})(\eta_{ij}^2 - \epsilon_{ij}^2) - \frac{\eta_{ij} + \epsilon_{ij}}{2}] = h(Y_{ij}|\beta, u_i) \tag{18}$$

Then the variational lower bound of the log of the likelihood for $\boldsymbol{Y}$is

$$\begin{aligned}
&\sum_{i=1}^{N} \log(\boldsymbol{Y}_i|\boldsymbol{\beta}, \sigma^2) \\
=&\sum_{i=1}^{N} \log \int P(Y_i|\beta, u_i)du_i \\
=&\sum_{i=1}^{N} \log \int \prod_{j=1}^{J_i} P(Y_{ij}|\beta, u_i)p(u_i|\sigma^2)du_i \\
\geq&\sum_{i=1}^{N} \log \int \prod_{j=1}^{J_i} h(Y_{ij}|\beta, u_i)p(u_i|\sigma^2)du_i
\end{aligned} \tag{19}$$

According to the form of $h(Y_{ij}|\beta, u_i)$ We conclude that the lower bound of the posterior of u_i is proportional to a normal distribution.

$$\prod_{j=1}^{J_i} h(Y_{ij}|\beta, u_i)p(u_i|\sigma^2) \propto N(u_i; \mu_i, \sigma_i^2), \tag{20}$$

where

$$\mu_i = \sigma_i^2 \sum_{j=1}^{J_i} (Y_{ij} - 2\lambda(\epsilon_{ij})x_{ij}^T\boldsymbol{\beta})z_{ij} \tag{21}$$

$$\sigma_i^2 = (\frac{1}{\sigma^2} + 2\sum_{j=1}^{J_i} \lambda(\epsilon_{ij})z_{ij}^2)^{-1} \tag{22}$$

The lower bound of the likelihood is

$$\begin{aligned}
&\sum_{i=1}^{N} \log(\boldsymbol{Y}_i|\boldsymbol{\beta}, \sigma^2) \\
=&\sum_{i=1}^{N} \log \int P(Y_i|\beta, u_i) du_i \\
\geq&\sum_{i=1}^{N} \log \int f(\boldsymbol{Y}_i, u_i|\boldsymbol{\beta}, \sigma^2) du_i \\
\geq&\sum_{i=1}^{N} \int q(u_i) \log \frac{f(\boldsymbol{Y}_i, u_i|\boldsymbol{\beta}, \sigma^2)}{q(u_i)} du_i \\
=&\mathbb{E}_q \log f(\boldsymbol{Y}_i, u_i|\boldsymbol{\beta}, \sigma^2) - \mathbb{E}_q \log q(u_i)
\end{aligned} \tag{23}$$

where$f(\boldsymbol{Y}_i, u_i|\boldsymbol{\beta}, \sigma^2) = \prod_{j=1}^{J_i} h(Y_{ij}|\beta, u_i) p(u_i|\sigma^2)$

Then we can use EM algorithms. In the E step, we get the posterior of u_i,

$$q(u_i|\boldsymbol{\beta}^{old}, \epsilon_{ij}^{old}, {\sigma^2}^{old}) = N(u_i; \mu_i, \sigma_i^2) \tag{24}$$

In the M step, we need to maximize $\mathbb{E}_q \log f(\boldsymbol{Y}_i, u_i|\boldsymbol{\beta})$

The updating equation of ϵ_{ij}^2 is

$$\epsilon_{ij}^2 = \left(x_{ij}^T \boldsymbol{\beta}^{old}\right)^2 + 2 x_{ij}^T \boldsymbol{\beta}^{old} z_{ij} u_i + z_{ij} (\mu_i^2 + \sigma_i^2) z_{ij}. \tag{25}$$

The updating equation of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta}^{old} = (\sum_{i=1}^{N} \sum_{j=1}^{J_i} \lambda(\epsilon_{ij}) x_{ij} x_{ij}^T)^{-1} (\sum_{i=1}^{N} \sum_{j=1}^{J_i} [Y_{ij} - 0.5 - 2\lambda(\epsilon_{ij}) z_{ij} \mu_i] x_{ij}) \tag{26}$$

The updating equation of σ is

$$\sigma^2 = \frac{1}{N} (\sum_{i=1}^{N} (\mu_i)^2 + \sigma_i^2) \tag{27}$$

This method uses the latent variables to considers more information of the data even though we can not observe these information from the data directly. But we still can get the improvement if the Signal to Noise Ratio(SNR) of the data is not good enough.

2.3 Neural Networks

We just use the Neural Networks as the introduction part. Neural Networks usually need much more data than common Machine Learning algorithms and are more computationally expensive than those traditional algorithms. If we have a massive amount of data and computational power that is available, We can consider the neural network. If we have a massive amount of data, the neural network may perform better than traditional programs.
If we dont have enough data or the computational power is not strong enough, a simple algorithm will be more suitable than the neural network.
In addition, the best-known disadvantage of Neural Networks is their black box nature. This means that its hard to understand which features are important and how they influence the responses. Interpretability is quite important in some domains, which means the neural network is not suitable.

The method is widely used and it would be great if the SNR of the data is good enough, like the image data. Compared with the linear regression, it needs more data and the calculation power due to its numerous parameters.

3 Experimental Study

3.1 Data Introduction

The sample size is over 40,000 and there are 47 dimensions of x, including weight, age and the feature of the PPG signal(like the derivative and the second derivative). And the y is $0, 1$.For $y = 0$, it means the sample is the normal. For $y = 1$, the sample is the hypertensive. The size of the test set is 0.3.

3.2 Experimental Result

3.2.1 Linear Logistic Regression

Our aim is to make the accuracy of the hypertensive group is over 85% and maximize the accuracy of the normal group.
For the Linear Logistic Regression, we can control t to adjust the accuracy of the hypertensive group and see the accuracy of the normal group. However, we get that the accuracy of hypertensive group is 83.48% and the normal group is 60.46%. And the ROC curve of the method is following:

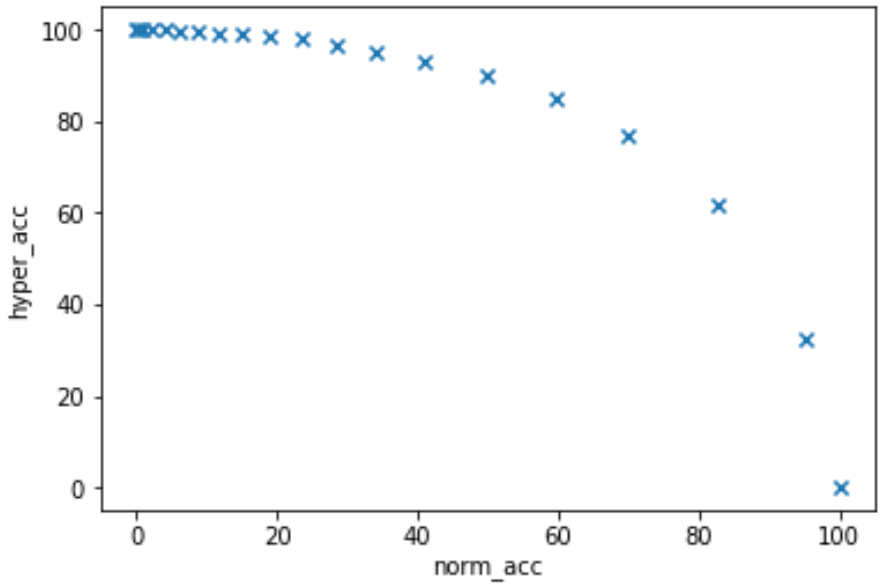

Figure 3: Linear Logistic Regression ROC Curve

The vertical axis is for the accuracy of the hypertensive group and the horizontal axis is for the accuracy of the normal group. This method can not satisfy our aim.

3.2.2 Variational Baysian Expectation Maximization

For the Variational Baysian Expectation Maximization, we can still control t to adjust the accuracy of the hypertensive group and see the accuracy of the normal group. It is really successful. We get that the accuracy of hypertensive group is 85.19% and the normal group is 84.81%. And the ROC curve of the method is following:

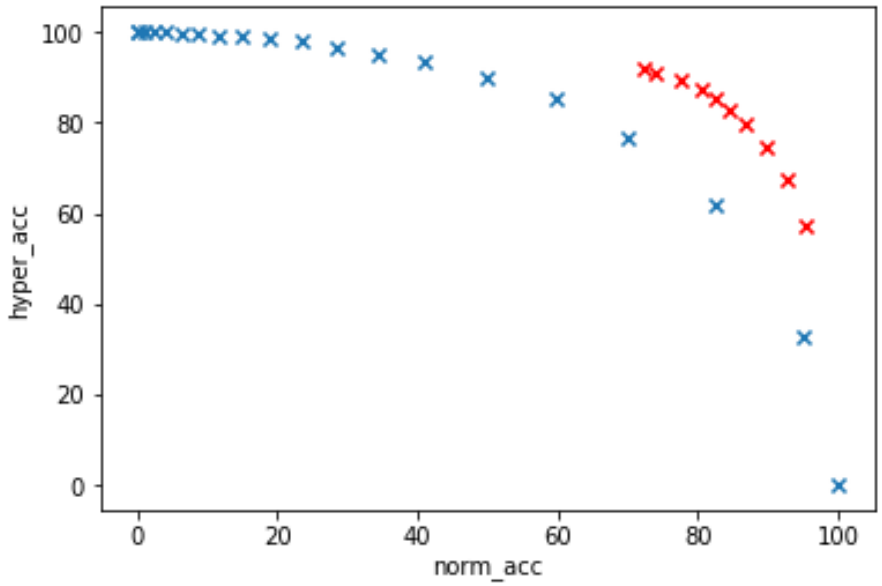

Figure 4: Variational Baysian Expectation Maximization ROC Curve

The red points comes from Variational Baysian Expectation Maximization. This method can use more information of the data even though the information do not show directly in the data. It shows its power and the advantage compared with Linear Logistic Regression

3.2.3 Neural Network

For the Neural Network,similarly, we can tune t to adjust the accuracy of the hypertensive group. However, it does not match our aim. The ROC curve of the method is following:

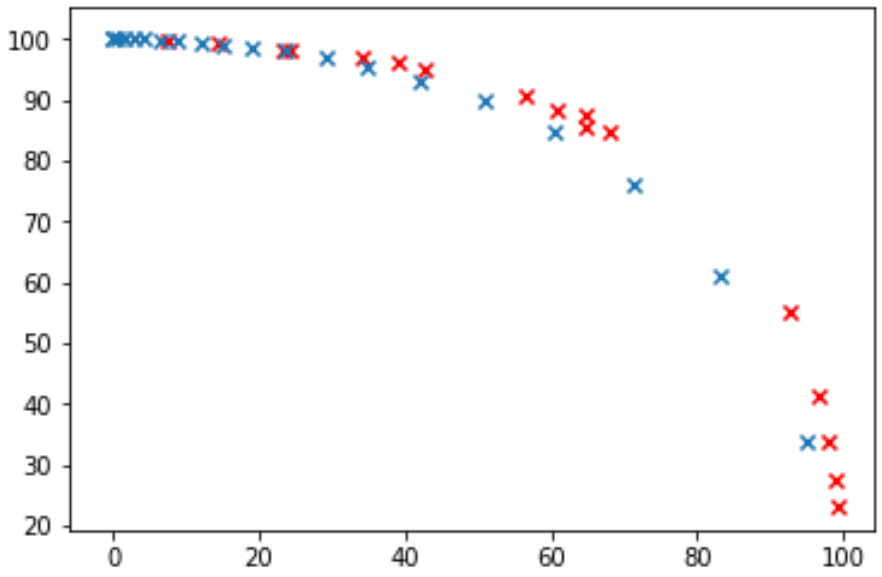

Figure 5: Neural Network ROC Curve

The red points comes from Neural Network. The results is a little bit better than Linear Logistic Regression but it can not satisfy our goal.

4 Conclusion

From the Experimental Study, we can find that Variational Bayesian Expectation Maximization is the best way for this classification problems and it satisfies our aims. The reason is that the data of this problems does not have the good SNR. At this condition, the Neural Network can be seen as the complex linear regression. So it is only a little bit better than the Logistic Linear Regression. However, Expectation Maximization can absorb more information if we can make a meaningful assumption based on the logic analysis. For example, we assume each person has his own random effect on the blood pressure. Due to this assumption, the Expectation Maximization can get more information and have a significant improvement.

This problem is really important. The percentage of hypertensives in China is over 40%, around 0.5 billion people. And some of them need to be monitored due to their Hypertensive complications. So continuous blood pressure measurement system without the blood pressure cuff is definitely

significant to those patients. It is the timely and accurate abnormal alarm which can safeguard the patients. In the future, more resources should be devoted to this research and more technologies should be applied in this problems. In the future, we may combine the neural network and EM algorithm to solve this problem.

5 Acknowledgements

I would like to express my gratitude to Prof Cheung Yiu Ming and TA Cai Wanling. They offer me a lot help during writing this paper. And my partner Chen Xiao, who provides the code of the Neural Network and we solve the Variational Baysian Expectation Maximization Algorithm together. The contribution: Lai Jianfa: 50%, Chen Xiao: 50%.

References

[1] Allen J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological measurement*, **28**(3), R1.

[2] B.M, Mallikarjun S. Holi, J. (2015). An Estimation Technique using FFT for Heart Rate Derived from PPG Signal. *Global Journal Of Research In Engineering*.

[3] Nasser M. Nasrabadi.(2007). Pattern Recognition and Machine Learning, *Journal of Electronic Imaging*, **16**(4).