WASHINGTON STATE
UNIVERSITY

Predicting Liver Disease Using a Feed Forward Neural Network

Aryan Ritwajeet Jha

Currently pursuing PhD ECE (Power Systems)

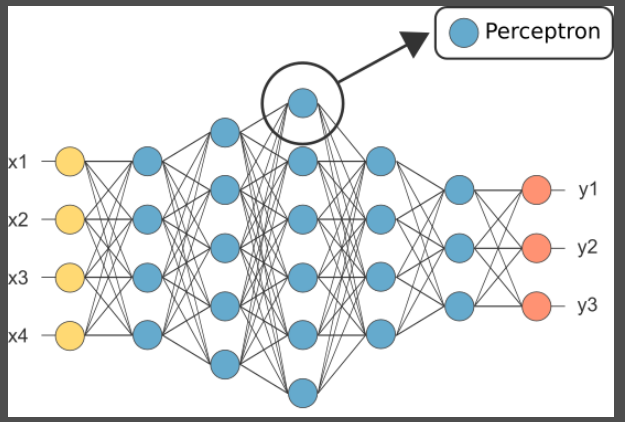
Problem Statement at a Glance [1, 2]

- Scenario: A patient arrives into the hospital with symptoms of liver disease
- Decision: To determine whether the patient has possible liver disease and be admitted for treatment
- Inputs: A set of 10 diagnostic quantities obtained from the patient.
- Additional Inputs for Training: A disease classification label for all arriving patients given by an expert physician.

Diagnostic Quantity	Type
Age	Integer
Gender	Nominal (Binary)
TB	Continuous
DB	Continuous
Alkphos	Integer
Sgpt	Integer
Sgot	Integer
TP	Continuous
ALB	Continuous
A/G Ratio	Continuous

Approach to Solve the Problem

- Train a Feed Forward Neural Type equation here.Network
- For 10 features and 1 classification, we can use a neural network represented by a vector [10, 10, 10, 1].
- For activation function, we use the sigmoid function.



Prediction Function

$$f(x) = \sigma\left(W_d\sigma\left(\cdots\sigma\left(W_2\sigma\left(W_1x\right)\right)\cdots\right)\right)$$

$$\begin{aligned} L_1 &= \sigma(W_1x) \\ L_2 &= \sigma(W_2L_1) \\ &\vdots \\ L_d &= \sigma(W_dL_{d-1}) \\ F &= \tfrac{1}{2}\left\|L_d - y\right\|^2 \end{aligned}$$

Backpropagation

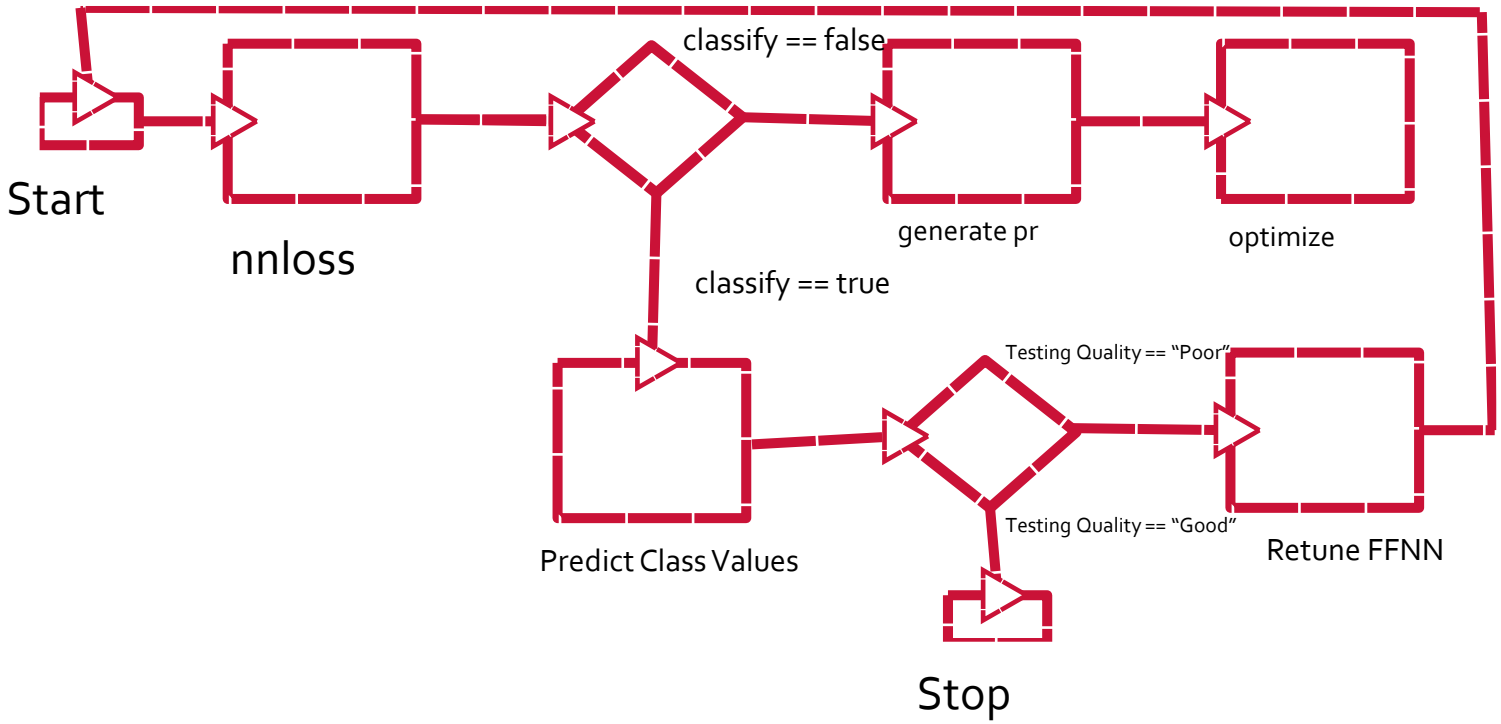
$$\begin{aligned} h_d &= (L_d - y)L_d(1 - L_d), \quad & G_d &= h_dL_{d-1}^{\mathsf{T}} \\ h_{d-1} &= W_d^{\mathsf{T}}h_dL_{d-1}(1 - L_{d-1}), \quad & G_{d-1} &= h_{d-1}L_{d-2}^{\mathsf{T}} \\ &\vdots \\ h_2 &= W_3^{\mathsf{T}}h_3L_2(1 - L_2), \quad & G_2 &= h_2L_1^{\mathsf{T}} \\ h_1 &= W_2^{\mathsf{T}}h_2L_1(1 - L_1), \quad & G_1 &= h_1x^{\mathsf{T}} \end{aligned}$$

Preprocessing

Given a chosen architecture for the FFNN, we would like to preprocess the dataset for our solver to have an easier time training the Neural Network.

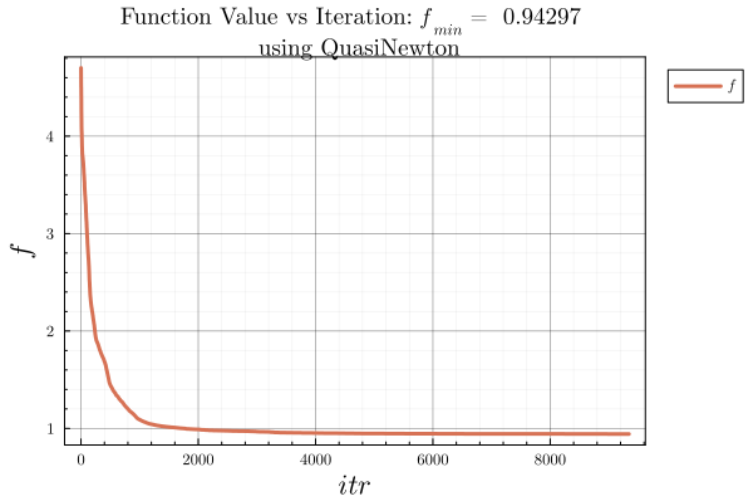
1. Drop Missing Data
2. Convert Nominal String values for Gender to Binary values. $\rightarrow \{0, 1\}$
3. Normalize all row values for a given column against the maximum value in that column $\rightarrow [0, 1]$
4. Choose training and testing data (Random $70 - 30$ split)
5. Choose initial weights $w0$
6. Map class values to $\{\frac{1}{3}, \frac{2}{3}\}$

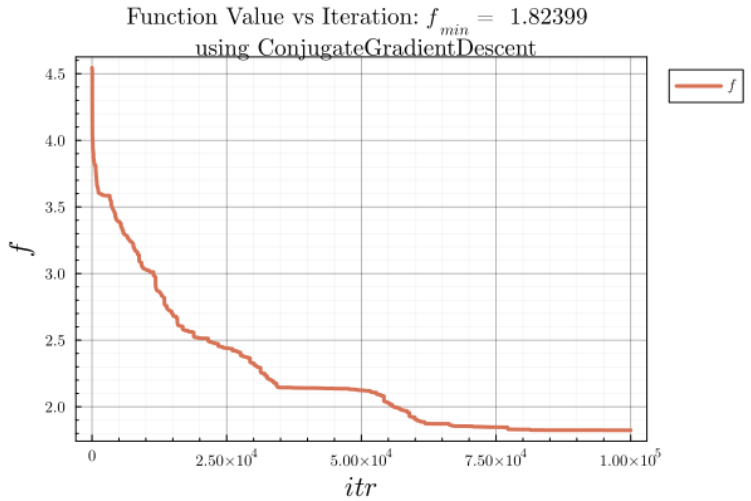
Code base for Training and Using the FFNN

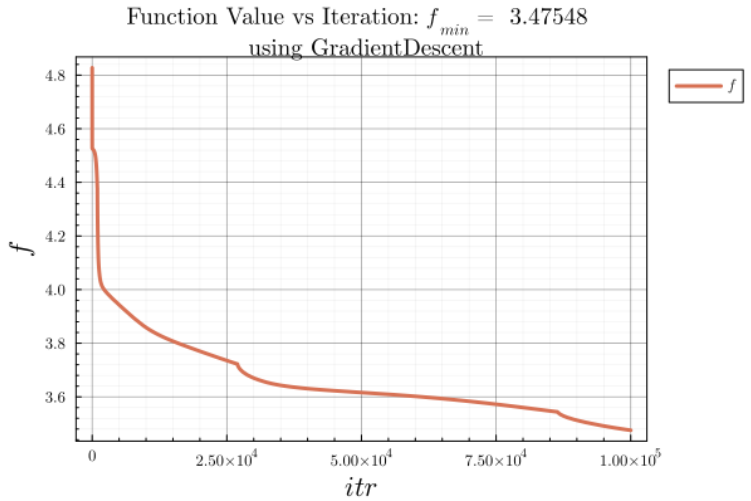


Some results

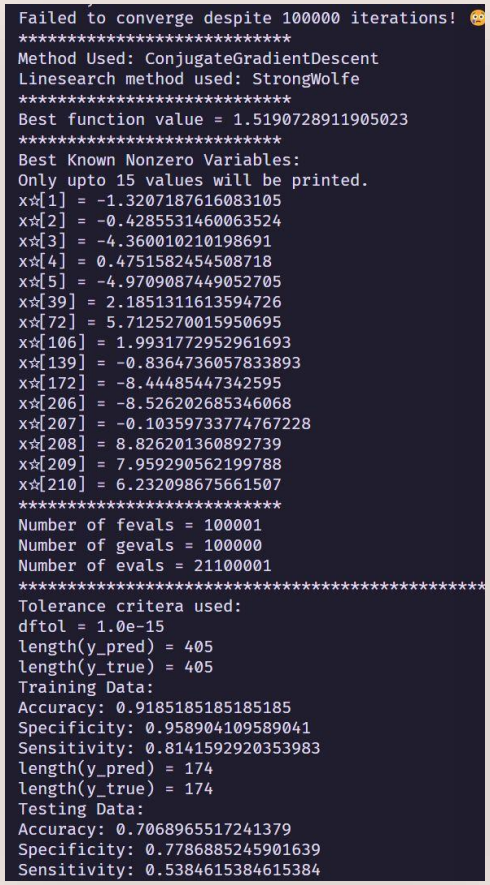
- Results varying, some better than others. Possible causes are:
  - Random splitting of training and testing datasets
  - Random initialization of weights
- For the 'base' FFNN of dimensions [10, 10, 10, 1]:

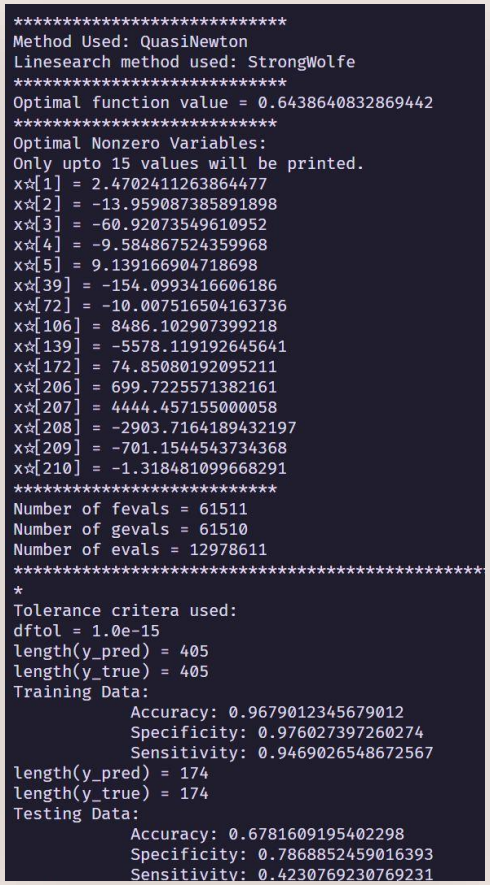


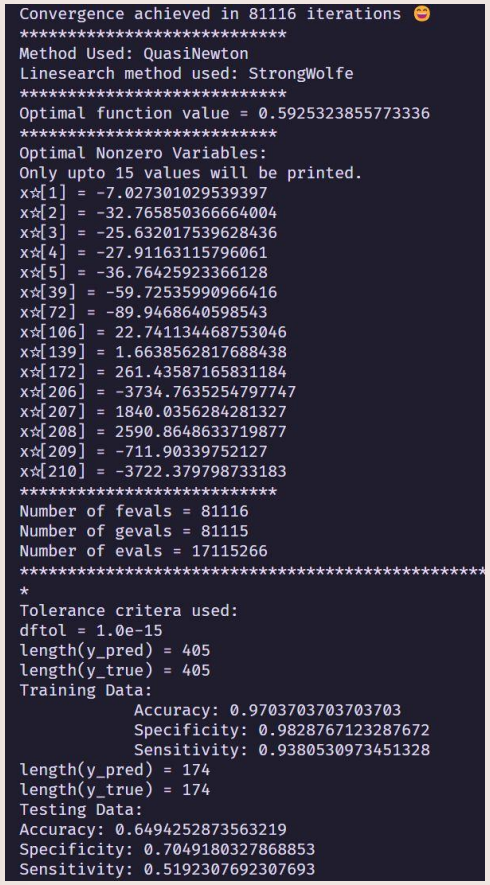




Some results



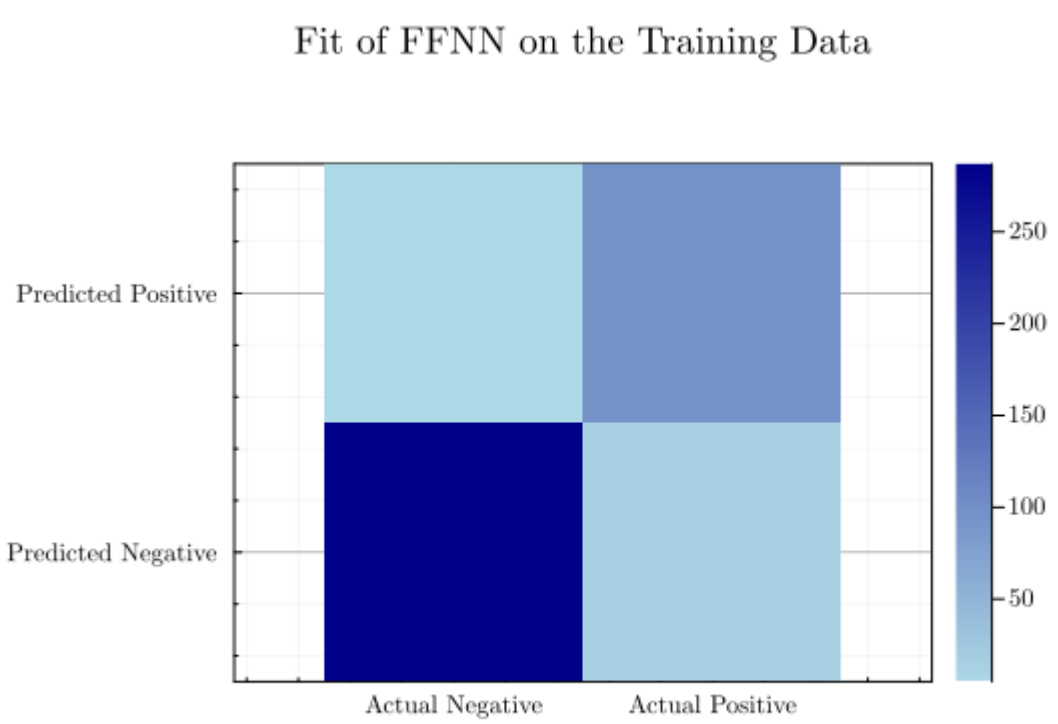


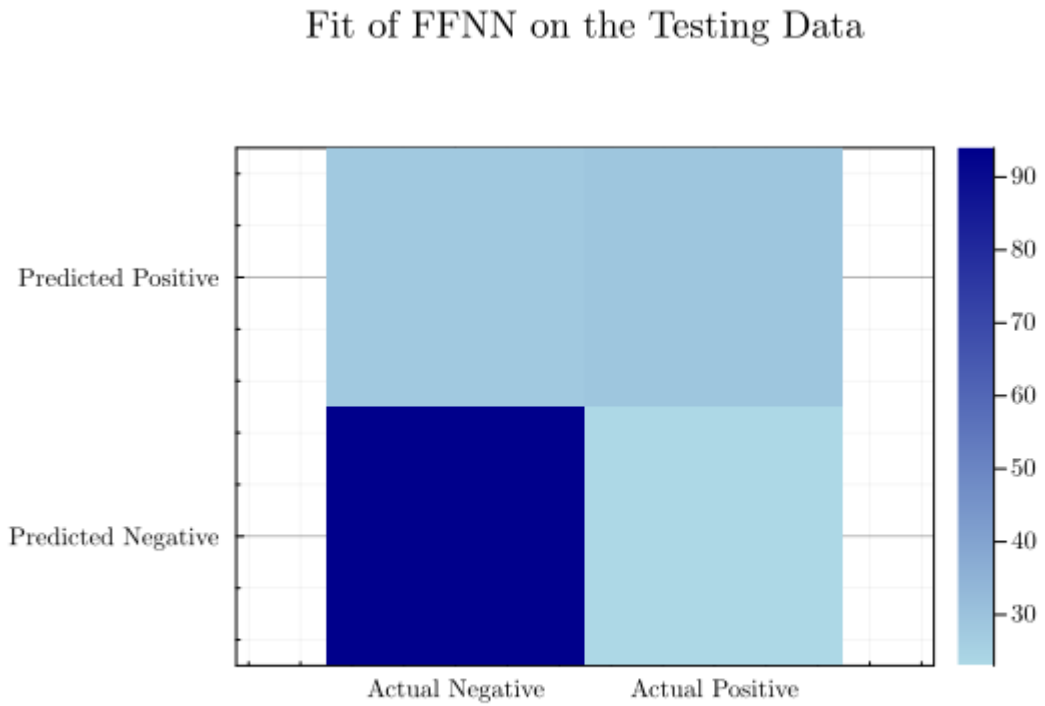


- Results varying, some better than others. Possible causes are:
- Training Accuracy varying between 91 – 97%
- Testing Accuracy varying between 65 – 71%
- For the ‘base’ FFNN of dimensions [10, 10, 10, 1]:

Some results

- Typical Confusion Matrix for the 'base' FFNN of dimensions [10, 10, 10, 1]:





Other NN architecture experimentation

(all runs using QN-BFGS)

- Using a single hidden layer: [10, 10, 1] : Poorer fit, even on the Training Data $(80 - 85\%)$
- Using less number of hidden nodes: $[10, 5, 5, 1]$: Less overfitting, gets lower Training Data accuracy (85%) but similar testing accuracy as 'base' architecture $(65 - 70\%)$.

Other untested ideas

- Try dropping some features. This paper [3] talks about how Albumin and A\G Ratio features are not that important when performing a diagnosis.

- Try another activation function, such as $\tanh()$ or $ReLU$.

Thank You.

Appendix and References follow

References

- [1] Ramana,Bendi and Venkateswarlu,N.. (2012). ILPD (Indian Liver Patient Dataset). UCI Machine Learning Repository. https://doi.org/10.24432/C5D02C.
- [2] Class notes and Problem Statements provided by Prof. Tom Asaki as part of the course MATH 564 Convex and Nonlinear Optimization at Washington State University, for the semester of Fall 2023.
- [3] Straw I, Wu | H Investigating for bias in healthcare algorithms: a sex-stratified analysis of supervised machine learning models in liver disease prediction | BMJ Health & Care Informatics 2022;29:e100457. doi: 10.1136/bmjhci-2021-100457