Правительство Российской Федерации

Федеральное государственное автономное образовательное учреждение

высшего образования

«Национальный исследовательский университет

“Высшая школа экономики”»

Жалкова Наталия Евгеньевна

**Интеллектуальная система диагностики здоровья человека**

Отчет по нейронной сети студента

2 курса бакалавриата группы НТ-6

Москва 2016

Вступление

На данный момент такая болезнь, как сахарный диабет, активно прогрессирует, число больных им людей заметно увеличивается с каждым годом – например, в США за 20 лет число больных диабетом увеличилось вдвое. Сейчас в развитых странах страдает диабетом от 5 до 12 % населения, впоследствии этот показатель вырастет до 30-35 %. В мире сейчас насчитывается около 360 млн. человек, больных диабетом.

Заболеть этой болезнью, к сожалению, может каждый человек и (что очень важно!) каждый уже может быть больным сахарным диабетом. Дело в том, что 85-90 % больных с диабетом так называемого 2 типа, а это – малосимптомное или вообще бессимптомное заболевание. Поэтому чтобы выявить диабет 2 типа, люди сами должны обратиться к врачу и сделать анализ – определить уровень глюкозы в крови.

К сожалению, данная болезнь не является такой безвредной, несмотря на почти отсутствие симптомов. При высоком содержании сахара в крови происходит разрушение сосудов глаз, почек, ног, сердца, головного мозга.

Сейчас сахарный диабет – частая причина слепоты, почечной недостаточности, так называемых нетравматических ампутаций. Но, несмотря на то, что это заболевание приводит к таким страшным осложнениям, это все можно предотвратить. Поэтому человек не должен ждать, пока появятся какие-то проблемы, надо своевременно и регулярно делать проверку глюкозы в крови.

Кроме того, делая данную проверку крови, можно выявить не только сахарный диабет, но и такие заболевания, как “Нарушение толерантности к глюкозе” – имеется риск развития сахарного диабета, а так же “Нарушенная гликемия натощак” – так же имеется склонность к диабету.

По приведенной ниже таблице можно легко понять границы, внутри которых возможно определить конкретное заболевание. На анализ берется или венозная или капиллярная кровь, которую необходимо сдать натощак. Так же для диагностики требуется кровь, взятая через 2 часа после приема внутрь растворенных в воде 75 граммов глюкозы.

Нормой для анализов являются: содержание глюкозы натощак меньше 5.6 ммоль/л, а через два часа – меньше 7.8 ммоль/л.



Обычно в Российских медицинских учреждениях у пациентов берут на анализ капиллярную кровь (из пальца). Поэтому в использованной мной выборке имеются данные именно об анализах капиллярной крови.

Я выбрала тему сахарного диабета из-за того, что данная проблема кажется мне серьезной и требующей внимания. С помощью внедрения нейросети станет проще получать диагноз, согласно полученным данным из анализа.

Анализ

У меня имеется 60 выборок результатов анализов различных пациентов, получивших свой диагноз. Среди них 5 с болезнью “Нарушенная гликемия натощак”, 6 людей с болезнью “Сахарный диабет”, 5 людей с выявленным “Нарушением толерантности к глюкозе” и 44 здоровых человека. Имеющиеся параметры, которые впоследствии будут передаваться в нейросеть как обучающие – состояние крови натощак и после двух часов с времени принятия глюкозы. Данные содержатся в табличке .xls, которая будет импортирована в MATLAB.

Для своей задачи я использую сеть так называемого типа Pattern recognition. Немного о ней:

Pattern recognition является сетью прямого распространения, то есть все связи направлены строго от входных нейронов к выходным и может быть обучена классификации входных данных относительно целей. В данном типе нейронной сети целями (выходными сигналами) должны быть векторы целиком из нулей, кроме единицы i-тым элементом, где i будет обозначать тот класс, к которому рассматривая выборка принадлежит. Таким образом, данный тип нейросети прекрасно подходит к решению моей задачи.

В качестве выходных сигналов, я возьму:

**1 0 0 0** – для сахарного диабета

**0 1 0 0** – для нарушения толерантности к глюкозе

**0 0 1 0** – для нарушения гликемии натощак

**0 0 0 1**– для нормы

В качестве входных – данные, содержащиеся в имеющейся базе выборок.

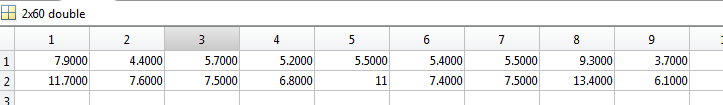
Таким образом, анализ завершен, подходящая для решения задачи нейросеть выбрана. Необходимо описать ее с помощью программы MATLAB.

Описание нейронной сети

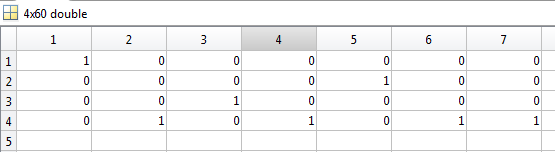
Привожу ниже выполненный мной порядок действий:

С помощью функции xlsread считываю данные из таблицы в переменную MATLAB.

>> A = xlsread('diabeth.xlsx');



Таким образом, получаю матрицу 2х60, содержащую результаты по анализам. Аналогичным методом записываю в переменную значения диагнозы, выявленные из анализов.



>> R = xlsread('results.xlsx');

Получаю матрицу 4х60. Данные записаны и готовы к обработке нейросетью. Создам нейросеть типа перцептрон. Для этого использую функцию patternnet. Ее синтаксис:

**patternnet(hiddenSizes, trainFcn, performFcn)**

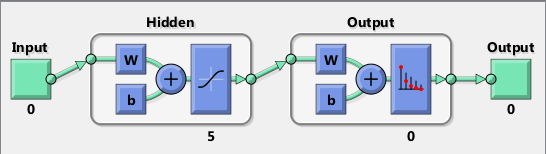
hiddenSizes – размеры скрытого слоя (по умолчанию = 10), trainFcn – функция, использующаяся в обучении (default = 'trainscg'), performFcn – функция оценки функционирования (default = 'crossentropy')

Соответственно:

>> net = patternnet(5);

Отображаю полученную нейросеть, чтобы увидеть ее строение.

>> view(net)

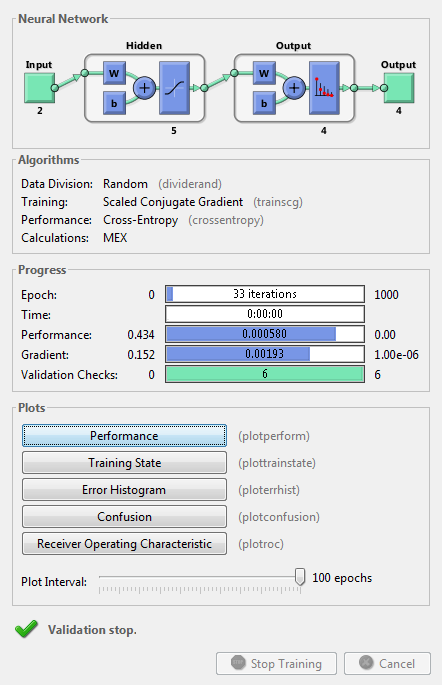


Размеры входа и выхода равны 0, т.к. нейросеть еще не была обучена, ей не были предоставлены никакие данные. Впоследствии это изменится. Так же на схеме можно увидеть скрытый слой, а так же внутренние функции.

Работа и анализ полученной сети

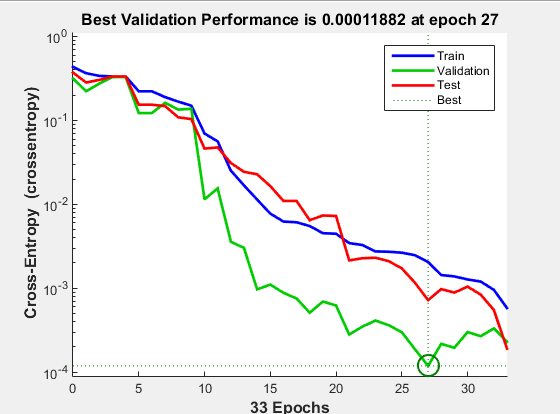
Начинаю обучение сети с помощью функции train. Выборка автоматически выбирает необходимые наборы “Train/Test/Validation”( обучающее 70%, тестирующее 25% и экзаменационное 5%). Для обучения используется набор обучающих данных. Обучение в данном случае продолжается до тех пор, пока нейросеть не прекращает улучшать экзаменационный набор данных. Тестовый же набор данных позволяет оценивать аккуратность работы нейросети.

>> [net, gr] = train(net, A, R);

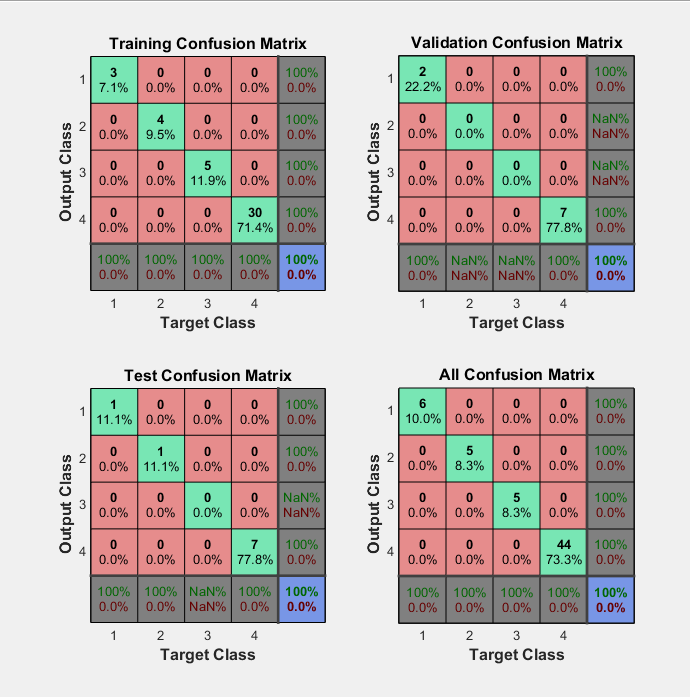


Дабы увидеть, как изменялась и улучшалась нейросеть во время обучения, нажимаю кнопку “Performance” или же вызвать команду plotperform(gr), при условии, что перед этим обучение нейросети было вызвано с помощью [net,gr] = train(net,x,t), где gr – запись обучения.

График отображается при этом в логарифмическом масштабе для того, чтобы видеть малейшие ошибки. С течением времени ошибка стремительно падает, пропорционально тому, как нейросеть обучается. На графике можно видеть все линии наборов, а так же момент времени, когда нейросеть добилась лучшего результата (выделено кружочком, пунктиром, а так же написано в заголовке).

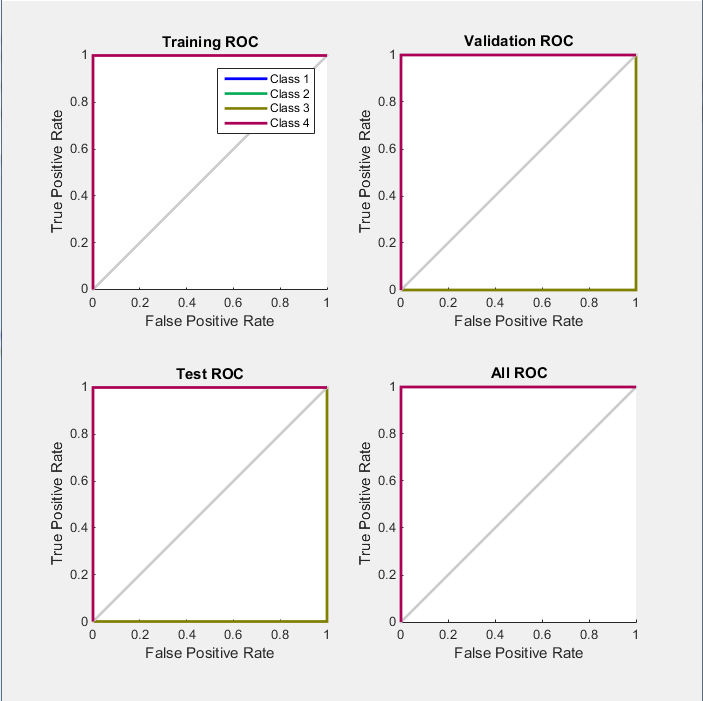


Еще одним показателем, насколько хорошо нейросеть была обучена введенным данным, является так называемый “confusion plot”, который можно вызвать, нажав кнопку Confusion. Матрица, которая появится в новом окне, показывает процент правильных и неправильных классификаций. Правильные классификации – зеленого цвета, неправильные – красного. Если нейросеть научилась правильно классифицировать вводимые данные, то проценты, указанные в красных квадратах, должны быть малы.

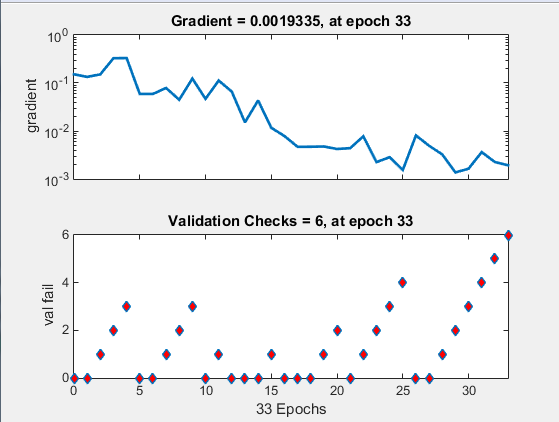


Можно наблюдать, что моя нейросеть показала замечательные результаты.

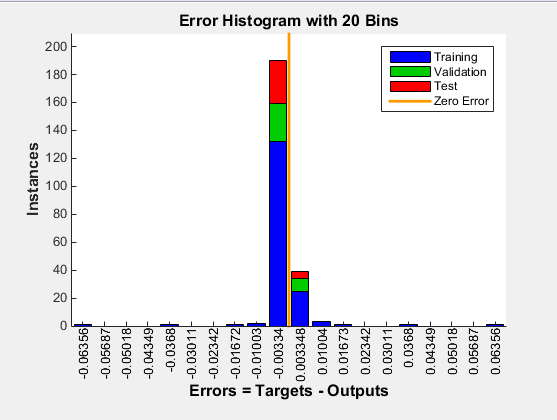
Еще одним показателем для нейросети является график, получаемый на кнопку “Receiver operating characteristic”. Данный график показывает, как зависят “false positive” и “true positive” оценки, на промежутке от 0 до 1. Как правило, в крайних углах (у единицы) можно наблюдать неровности, обусловленные неточностями. В созданной мной нейронной сети данная ошибка близка к нулю.



Кроме того, результат обучения нейросети можно увидеть через кнопку “Training State”, где можно увидеть изменение градиента во время обучения, а так же сколько неудач было совершенно в конкретную эпоху при сравнении обучающего подмножества и экзаменационного. 6 = это максимально допустимый уровень превышения ошибки контрольного подмножества по сравнению с обучающим.



Осталась последняя кнопка “Error histogram”, которая отображает гистограмму ошибок = цели – выход.



Кроме того, всю полученную информацию можно увидеть, вызвав gr, объявленную при обучении нейросети. Это структура, которая содержит информацию о проведенном обучении.

>> gr

gr =

trainFcn: 'trainscg'

trainParam: [1x1 struct]

performFcn: 'crossentropy'

performParam: [1x1 struct]

derivFcn: 'defaultderiv'

divideFcn: 'dividerand'

divideMode: 'sample'

divideParam: [1x1 struct]

trainInd: [1x42 double]

valInd: [1 13 18 27 32 33 43 49 51]

testInd: [15 19 20 22 24 30 41 42 45]

stop: 'Validation stop.'

num\_epochs: 33

trainMask: {[4x60 double]}

valMask: {[4x60 double]}

testMask: {[4x60 double]}

best\_epoch: 27

goal: 0

states: {'epoch' 'time' 'perf' 'vperf' 'tperf' 'gradient' 'val\_fail'}

epoch: [1x34 double]

time: [1x34 double]

perf: [1x34 double]

vperf: [1x34 double]

tperf: [1x34 double]

gradient: [1x34 double]

val\_fail: [0 0 1 2 3 0 0 1 2 3 0 1 0 0 0 1 0 0 0 1 2 0 1 2 3 4 0 0 1 2 3 4 5 6]

best\_perf: 0.0020

best\_vperf: 1.1882e-04

best\_tperf: 7.2194e-04

Результаты работы, тесты

Теперь полученную нейросеть можно протестировать новыми наборами данных, которые не были использованы во время обучения. Данную операцию стоит проводить для того, чтобы узнать, насколько хорошо нейросеть может определять сама диагноз. Для этого воспользуюсь

>> testing = net([7; 13]); - передача нейросети входного вектора с данными.

answer = testing > 0.5 – перевод в логическую переменную (вид 1 0 0 0)

Так же можно вызвать функцию y = sim(net,[7; 13])

Привожу таблицу с проведенными тестами:

|  |  |  |
| --- | --- | --- |
| Вводимые данные | | Ответ нейросети |
| Количество глюкозы  натощак (ммоль/л) | Количество глюкозы  Через 2 часа (ммоль/л) | Диагноз |
| 7 | 13 | 1000 - диабет |
| 5 | 6 | 0001 - норма |
| 5.6 | 9.4 | 0100 – наруш. толер. |
| 6 | 5.4 | 0010 – наруш. глик. |
| 5.2 | 6.7 | 0001 - норма |
| 5.8 | 5.1 | 0001 - норма |
| 8.7 | 15.9 | 1000 - диабет |
| 6.6 | 11.7 | 1000 - диабет |
| 6 | 8.9 | 0100 - наруш. толер. |
| 6.1 | 5.7 | 0010 – наруш. глик. |
| 7.7 | 10.5 | 1000 - диабет |
| 4.3 | 8.7 | 0001 - норма |
| 5.8 | 7.2 | 0010 – наруш. глик. |

Выводы

Результаты показали, что нейросеть хорошо обучена, но, к сожалению, встречаются редкие ошибки. Они почти все относятся к диагностированию нарушенной гликемии натощак, что вызвано маленькими границами для данных и сильной близостью к нормальному состоянию. Но даже несмотря на это, нейросеть умеет распознавать и этот тонкий результат. Подобные ошибки можно убрать добавлением количества выборок (вводимых данных+диагноз). Наиболее точно определяется сахарный диабет, что, несомненно, большой плюс для нейросети, разработанной для этой задачи.