МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

«Вятский государственный университет»

Факультет автоматики и вычислительной техники

Кафедра ЭВМ

Отчёт

Лабораторная работа № 2 по дисциплине

«Системы обработки знаний»

Выполнил студент группы ИВТб-4301\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ / Жеребцов К. А./

Проверил доцента кафедры ЭВМ\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_/ Ростовцев В. С./

Киров 2024

1. Цель

Ознакомиться с основными командами создания, обучения и применения нейронных сетей кластеризации и классификации данных в Neural Network Toolbox с помощью сетей Кохонена и LVQ.

1. Задание

1) Выполнить кластеризации с применением нейронных сетей Кохонена в соответствие с заданием (Приложение А). Создать и выполнить моделирование нейронной Кохонена, согласно согласованному с преподавателем, приведенному в приложении А. Результаты кластеризации вывести в виде графика с разными цветами кластеров, а также точки тестовых векторов.

2) Для создания сети LVQ обучающую последовательность использовать сеть Кохонена.

1. Протокол выполнения
   1. Самоорганизующаяся карта Кохонена
      1. Обучение

Для кластеризации данных был найдена датасет, состоящий из массы и роста людей. Датасет приведен в приложении А.

Для отображения элементов выборки был выполнен следующий код в Matlab:

plot(indata(1,:),indata(2,:),'.g','markersize',20)

hold on

Результат выполнения команд представлен на рисунке 1.

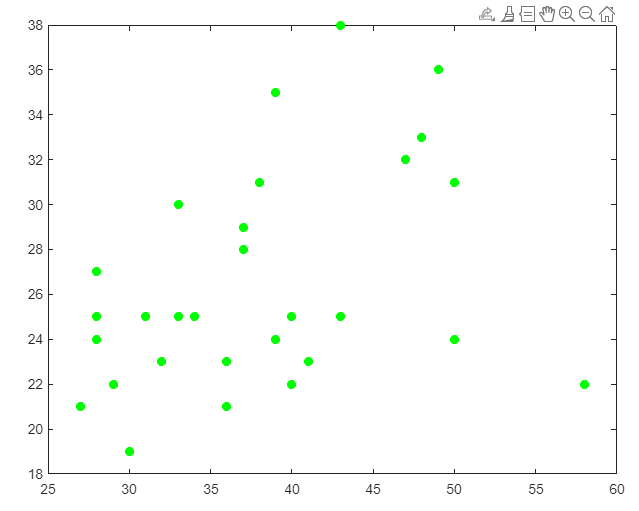


Рисунок 1 – Отображение элементов обучающей выборки

Из-за того, что в обучающей выборке есть похожие данные, то точки накладываются друг на друга.

Для создания самоорганизующейся карты Кохонена выл выполнен следующий код в программе MatLab:

net = selforgmap([6,6]);

Была создана гексагональная сетка размера 6х6, что соответствует 36 нейронам или, условно, кластерам.

Для обучения сети использовалась следующая команда:

[net, tr] = train(net, indata);

Процесс обучения изображен на рисунке 2.

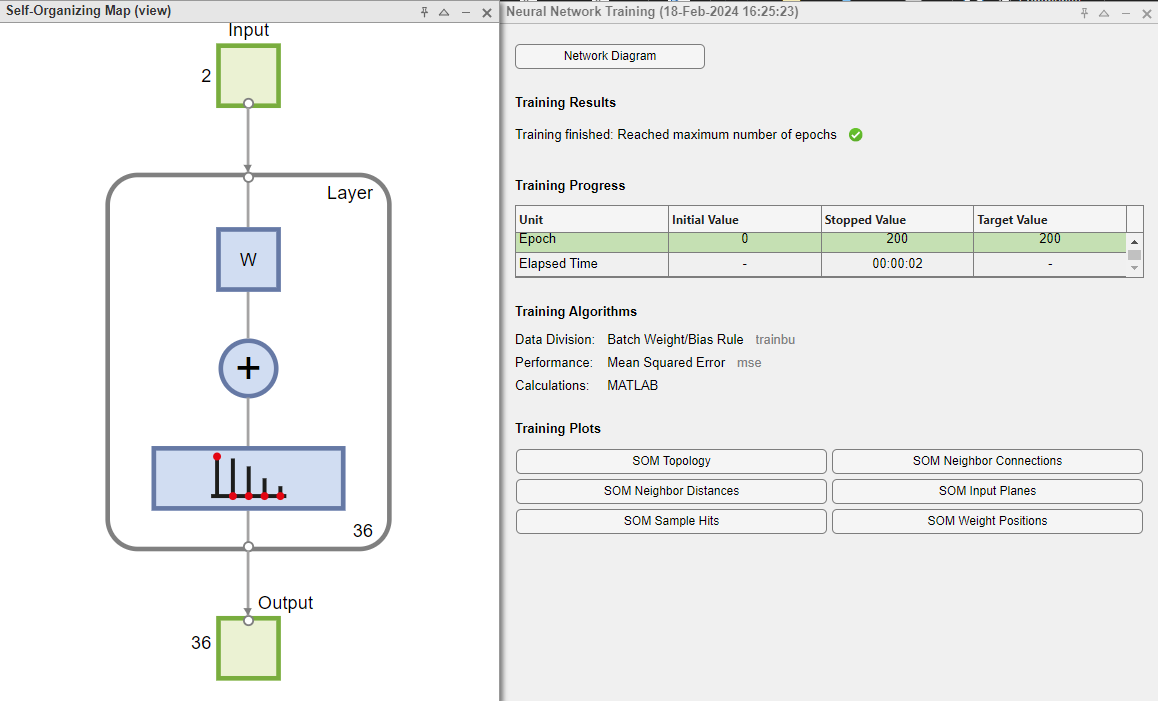


Рисунок 2 – Процесс обучения сети

В результате были получены новые значения весов нейронов. Узлы карты Кохонена были отображена на графике с помощью следующего кода:

W = net.iw{1,1}

for i = 1:length(W)

text(W(i,1),W(i,2),int2str(i),'HorizontalAlignment','center', 'VerticalAlignment', 'bottom')

end

hold on

plotsom(net.iw{1,1},net.layers{1}.distances)

hold off

Результат выполнения кода представлен на рисунке 2.

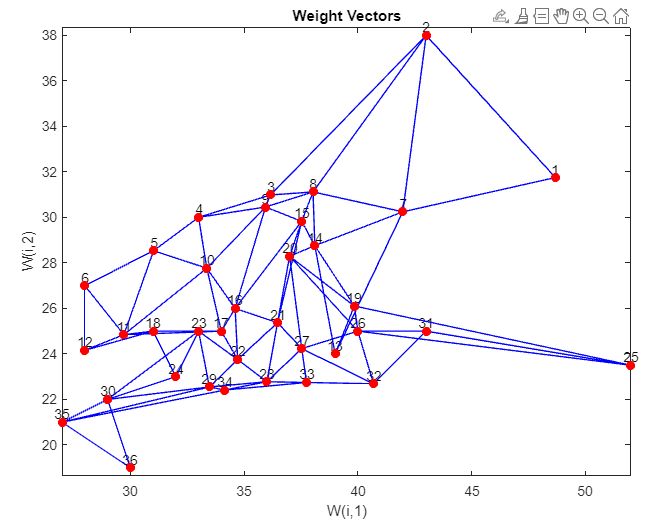


Рисунок 3 – Отображение узлов карты Кохонена после обучения

Как видно из полученного графика, были выделены кластеры, размером с заданной картой. Особо выделаются крайние кластеры, которые соответствуют группам людей с редким возрастом или ИМТ.

* + 1. Тестирование сети Кохонена

Для тестирования обучения сети выделим несколько примеров, которые не участвовали в обучении:

* Человек возрастом 53 и ИМТ 25.
* Человек возрастом 39 и ИМТ 35.
* Человек возрастом 28 и ИМТ 24.

Отобразим их на графике синим цветом, рисунок 4.

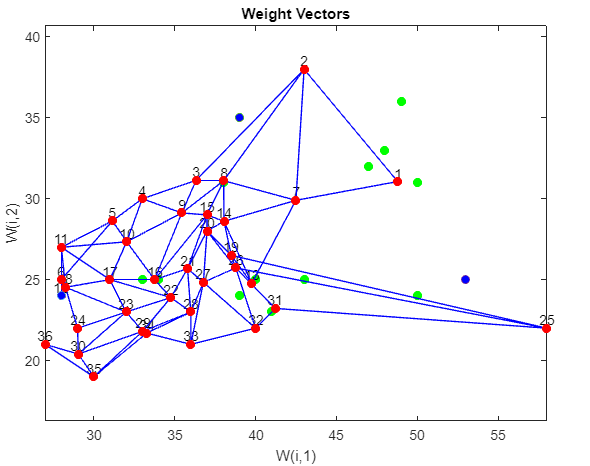


Рисунок 4 – Отображение тестовых примеров на графике синим цветом

Выполним следующий код:

res = sim(net, [53; 25])

for i = 1:length(res)

disp([num2str(i) ': ' num2str(res(i))])

end

Вывод представлен на рисунке 5.



Рисунок 5 – Тестирование принадлежности человека возрастом 53 и ИМТ 25 к кластеру 25.

Как видно из рисунка 5, тестирование прошло успешно – человек возрастом 53 и ИМТ 25 был отнесен в группу 25, что соответствует графике, где он близок к узлу под номером 25.

Проделаем аналогичные действия для человека возрастом 39 и ИМТ 35, а также возрастом 28 и ИМТ 24, рисунки 6 и 7.



Рисунок 6 – Тестирование человека возрастом 39 и ИМТ 35.

На графике видно, что человек возрастом 39 и ИМТ 35 близок к кластерам 3 и 8. Но ближе всего он оказался к узлу под номером 8 (рисунок 6), что соответствует графику на рисунке 4.



Рисунок 7 – Тестирование человек возрастом 28 и ИМТ 24.

Судя по рисунку 7, человек возрастом 28 и ИМТ 24 соответствует группе 12, что полностью соответствует рисунку 4.

* + 1. Выводы по сети Кохонена

В данном разделе была создана сеть Кохонена в программе Matlab. Для неё была найдена обучающая выборка, состоящая из возраста и ИМТ людей. Сеть была обучена, в результате чего экземпляры выборки были отнесены к определенным кластерам, что было показано на графике. Для тестирования использовались 3 примера, которые не участвовали в обучении сети.

* 1. Сеть LVQ
     1. Создание и обучение

Для создания сети LVQ возьмем обучающую последовательность, используя результаты сети Кохонена.

Разобьем людей на 9 классов с помощью сети Кохонена. Результат показан на рисунке 8.

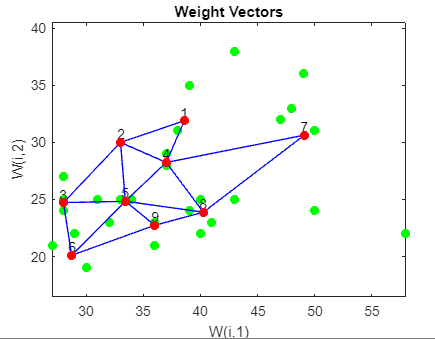


Рисунок 8 – Разбиение людей на 9 классов

Заменим точки числами их классов, также зададим им разный цвет следующим кодом:

T = []

for i = 1:length(indata)

y = net([indata(1, i); indata(2, i)]);

cluster\_index = vec2ind(y);

T = [T cluster\_index];

end

colors = ['r', 'r', 'r', 'r', 'b','b','b','b','g','g','g','g','c','c','c','c']

figure(1), clf, axis([min(indata(1,:))-5,max(indata(1,:))+5,min(indata(2,:))-5,max(indata(2,:)+5)]), hold on

for i = 1:12

tmp = find(T==i);

s = num2str(i) + 'r'

% plot(indata(1,tmp),indata(2,tmp), s)

text(indata(1,tmp),indata(2,tmp), num2str(i),'Color',colors(i))

hold on

end

Результат представлен на рисунке 9.

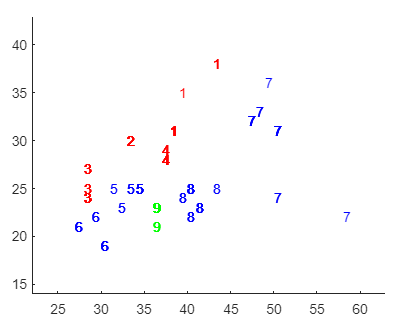


Рисунок 9 – Визуальное представление классов

Создадим сеть LVQ и обучим её:

t = ind2vec(T);

lnet = lvqnet(18,0.2,'learnlv2');

lnet.trainParam.epochs=150;

lnet=train(lnet,indata,t);

Процесс обучения представлен на рисунке 10.

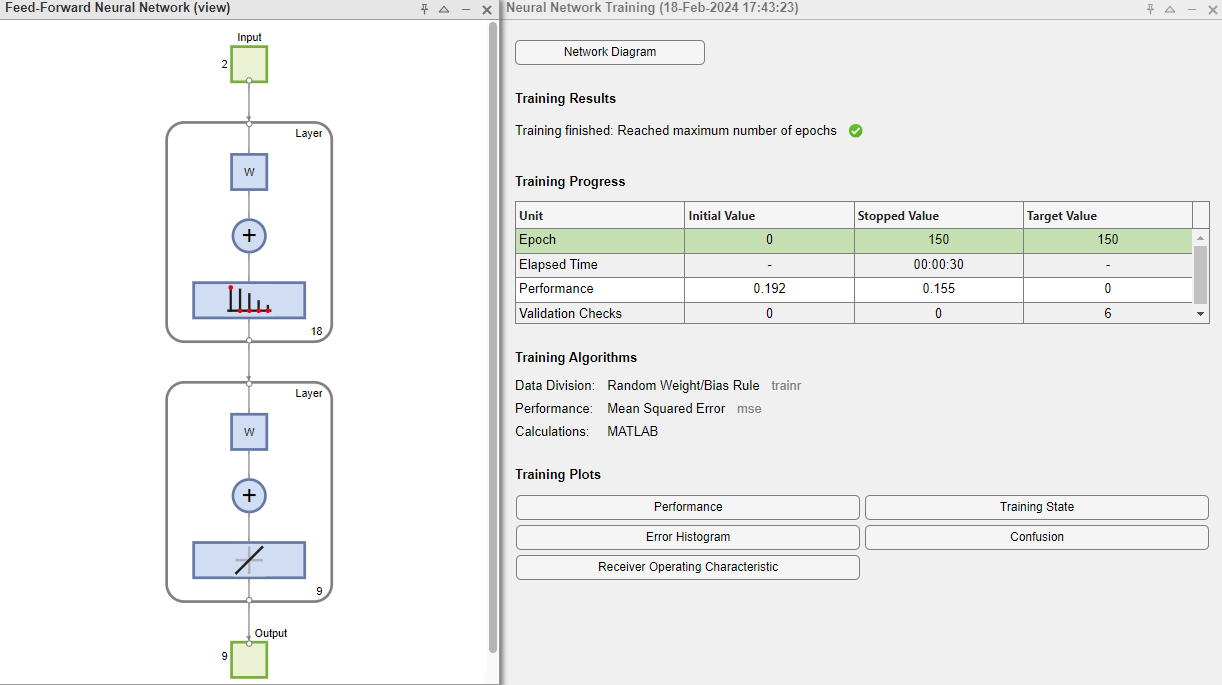
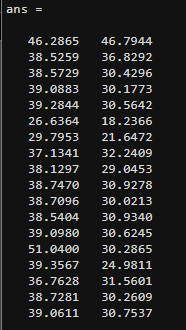


Рисунок 10 – Процесс обучения сети LVQ

После обучения получаем следующие весовые коэффициенты нейронов конкурирующего слоя, которые определяют положения центров кластеризации:



Отобразим их на графике (рисунок 11, цвет точек центров не зависит от цветов изначальных кластеров):

plotvec(lnet.IW{1}',vec2ind(lnet.LW{2}),'o');

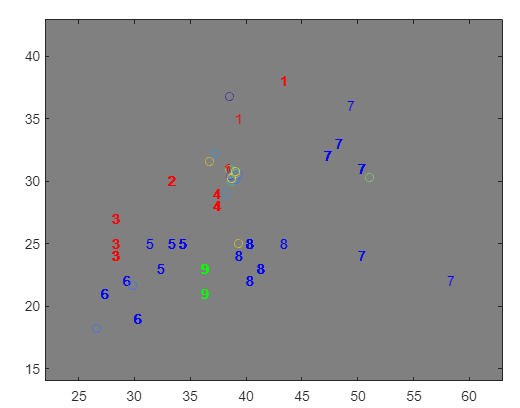


Рисунок 11 – Центры кластеризации

Наибольшее скопление кластеров произошло на стыке 1 и 4 групп.

* + 1. Тестирование сети

Для тестирования сети возьмем 3 примера:

* Человека возрастом 58 и ИМТ 22.
* Человек возрастом 25 и ИМТ 31.
* Человек возрастом 39 и ИМТ 27.

Тестирование показано на рисунке 12.

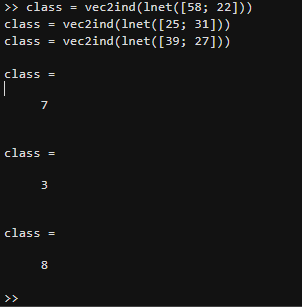


Рисунок 12 – Тестирование сети

Как видно из тестирования, есть небольшие смещения центров кластеров относительно изначальных групп, поэтому ответы немного не совпадают с их классами. Например, Человек возрастом 39 и ИМТ 27 по факту принадлежит 4 классу, но тестирование показало 8.

Человек возрастом 25 и ИМТ 31 отнесен к 3 классу, что полностью соответствует графику 11.

Человек возрастом 58 и ИМТ 22 отнесен к 7 классу, что полностью соответствует графику 11.

* + 1. Выводы по сети LVQ

В данном разделе была создана сеть LVQ в программе Matlab. В отличие от сети Кохонена, сеть LVQ требует разбиение входных векторов на классы, поэтому для обучения сети были взяты результаты обучения сети Кохонена. Сеть была обучена, в результате чего экземпляры выборки были отнесены к определенным кластерам, что было показано на графике. Для тестирования использовались 3 примера, которые не участвовали в обучении сети.

1. Выводы

В данной лабораторной работе были изучены основные команды создания, обучения и применения нейронных сетей кластеризации и классификации данных в Neural Network Toolbox с помощью сетей Кохонена и LVQ.

Основная идея кластерного анализа (clustering, cluster analysis) заключается в том, чтобы разбить объекты на группы или кластеры таким образом, чтобы внутри группы эти наблюдения были более похожи друг на друга, чем на объекты другого кластера.

Алгоритм функционирования самообучающихся карт (Self Organizing Maps — SOM) представляет собой один из вариантов кластеризации многомерных векторов. Примером таких алгоритмов может служить алгоритм k-ближайших средних (k-means). Важным отличием алгоритма SOM является то, что в нем все нейроны (узлы, центры классов…) упорядочены в некоторую структуру (обычно двумерную сетку).

При этом в ходе обучения модифицируется не только нейрон-победитель, но и его соседи, но в меньшей степени. За счет этого SOM можно считать одним из методов проецирования многомерного пространства в пространство с более низкой размерностью. При использовании этого алгоритма вектора, схожие в исходном пространстве, оказываются рядом и на полученной карте.

LVQ-сети основаны на обучающемся векторном квантовании (LVQ — Learning Vector Quantization) и представляют собой слой Кохонена, обучающийся с учителем. Для построения LVQ-сети задается количество кластеров (нейронов) n, количество классов m ( n ³ m ) и принадлежность каждого кластера определенному классу. Например, анализы пациентов разделяются на 5 кластеров, 2 из которых соответствуют здоровым людям, а 3 — больным. Разделить кластеры по классам можно в тех же пропорциях, что и распределение примеров соответствующих классов в обучающей выборке. Соответствие между номерами кластеров и номерами классов может быть произвольным. Для простоты интерпретации результатов работы сети целесообразно пронумеровать кластеры последовательно. В процессе обучения LVQ-сети веса нейронов настраиваются с учетом принадлежности обучающих примеров и кластеров одному классу. Обученная LVQ-сеть производит кластеризацию входных векторов с учетом классов.

Приложение А

Датасет для самоорганизующейся карты Кохонена. Данные взяты с сайта: <https://archive.ics.uci.edu/dataset/445/absenteeism+at+work>

|  |  |
| --- | --- |
| Age | BMI |
| 33 | 30 |
| 50 | 31 |
| 38 | 31 |
| 39 | 24 |
| 33 | 30 |
| 38 | 31 |
| 28 | 27 |
| 36 | 23 |
| 34 | 25 |
| 37 | 29 |
| 36 | 23 |
| 36 | 23 |
| 36 | 23 |
| 38 | 31 |
| 38 | 31 |
| 41 | 23 |
| 38 | 31 |
| 38 | 31 |
| 33 | 25 |
| 47 | 32 |
| 28 | 25 |
| 38 | 31 |
| 28 | 27 |
| 36 | 23 |
| 33 | 30 |
| 28 | 27 |
| 33 | 30 |
| 29 | 22 |
| 33 | 30 |
| 38 | 31 |
| 38 | 31 |
| 48 | 33 |
| 37 | 29 |
| 48 | 33 |
| 38 | 31 |
| 28 | 27 |
| 33 | 30 |
| 32 | 23 |
| 48 | 33 |
| 36 | 23 |
| 27 | 21 |
| 37 | 28 |
| 38 | 31 |
| 43 | 38 |
| 34 | 25 |
| 37 | 28 |
| 38 | 31 |
| 40 | 25 |
| 36 | 23 |
| 40 | 25 |
| 36 | 23 |
| 28 | 24 |
| 28 | 24 |
| 37 | 28 |
| 33 | 30 |
| 50 | 31 |
| 28 | 24 |
| 38 | 31 |
| 31 | 25 |
| 47 | 32 |
| 38 | 31 |
| 36 | 23 |
| 38 | 31 |
| 37 | 28 |
| 50 | 31 |
| 30 | 19 |
| 38 | 31 |
| 28 | 24 |
| 37 | 28 |
| 28 | 24 |
| 47 | 32 |
| 40 | 25 |
| 38 | 31 |
| 28 | 24 |
| 36 | 23 |
| 40 | 25 |
| 28 | 24 |
| 33 | 30 |
| 28 | 27 |
| 36 | 23 |
| 38 | 31 |
| 28 | 24 |
| 38 | 31 |
| 40 | 22 |
| 40 | 25 |
| 34 | 25 |
| 33 | 25 |
| 40 | 25 |
| 28 | 24 |
| 34 | 25 |
| 28 | 24 |
| 40 | 22 |
| 28 | 24 |
| 36 | 23 |
| 47 | 32 |
| 28 | 24 |
| 33 | 30 |
| 40 | 25 |
| 47 | 32 |
| 37 | 28 |
| 50 | 31 |
| 37 | 29 |
| 28 | 24 |
| 36 | 23 |
| 37 | 28 |
| 28 | 27 |
| 28 | 24 |
| 36 | 23 |
| 28 | 24 |
| 28 | 27 |
| 37 | 28 |
| 41 | 23 |
| 28 | 24 |
| 28 | 24 |
| 37 | 28 |
| 37 | 28 |
| 34 | 25 |
| 28 | 24 |
| 27 | 21 |
| 28 | 24 |
| 28 | 24 |
| 37 | 28 |
| 28 | 24 |
| 37 | 28 |
| 37 | 28 |
| 37 | 28 |
| 37 | 28 |
| 37 | 28 |
| 37 | 28 |
| 30 | 19 |
| 33 | 30 |
| 37 | 28 |
| 27 | 21 |
| 37 | 28 |
| 37 | 28 |
| 28 | 24 |
| 33 | 30 |
| 27 | 21 |
| 41 | 23 |
| 38 | 31 |
| 34 | 25 |
| 33 | 25 |
| 36 | 23 |
| 33 | 30 |
| 50 | 24 |
| 50 | 24 |
| 28 | 24 |
| 28 | 24 |
| 30 | 19 |
| 27 | 21 |
| 28 | 24 |
| 28 | 25 |
| 28 | 25 |
| 28 | 24 |
| 33 | 25 |
| 32 | 23 |
| 36 | 23 |
| 29 | 22 |
| 40 | 22 |
| 40 | 25 |
| 36 | 23 |
| 30 | 19 |
| 47 | 32 |
| 36 | 23 |
| 40 | 22 |
| 34 | 25 |
| 36 | 23 |
| 34 | 25 |
| 33 | 30 |
| 40 | 22 |
| 36 | 23 |
| 28 | 24 |
| 39 | 24 |
| 38 | 31 |
| 28 | 24 |
| 28 | 24 |
| 30 | 19 |
| 28 | 24 |
| 28 | 24 |
| 38 | 31 |
| 39 | 24 |
| 28 | 24 |
| 47 | 32 |
| 28 | 24 |
| 40 | 25 |
| 28 | 24 |
| 34 | 25 |
| 41 | 23 |
| 34 | 25 |
| 28 | 24 |
| 36 | 23 |
| 38 | 31 |
| 50 | 31 |
| 40 | 25 |
| 41 | 23 |
| 40 | 25 |
| 47 | 32 |
| 36 | 23 |
| 33 | 30 |
| 34 | 25 |
| 36 | 21 |
| 33 | 30 |
| 37 | 29 |
| 48 | 33 |
| 33 | 30 |
| 34 | 25 |
| 34 | 25 |
| 38 | 31 |
| 28 | 24 |
| 27 | 21 |
| 34 | 25 |
| 38 | 31 |
| 33 | 30 |
| 39 | 24 |
| 28 | 25 |
| 36 | 21 |
| 50 | 24 |
| 38 | 31 |
| 50 | 31 |
| 28 | 27 |
| 41 | 23 |
| 28 | 27 |
| 41 | 23 |
| 40 | 25 |
| 41 | 23 |
| 38 | 31 |
| 34 | 25 |
| 41 | 23 |
| 50 | 31 |
| 37 | 29 |
| 50 | 31 |
| 50 | 31 |
| 36 | 21 |
| 38 | 31 |
| 49 | 36 |
| 28 | 24 |
| 34 | 25 |
| 50 | 31 |
| 37 | 28 |
| 37 | 28 |
| 28 | 25 |
| 30 | 19 |
| 34 | 25 |
| 28 | 25 |
| 28 | 25 |
| 29 | 22 |
| 28 | 27 |
| 33 | 30 |
| 38 | 31 |
| 33 | 30 |
| 33 | 30 |
| 36 | 23 |
| 33 | 30 |
| 29 | 22 |
| 33 | 30 |
| 58 | 22 |
| 43 | 25 |
| 43 | 25 |
| 36 | 23 |
| 33 | 30 |
| 47 | 32 |
| 33 | 25 |
| 30 | 19 |
| 50 | 31 |
| 47 | 32 |
| 37 | 29 |
| 50 | 31 |
| 37 | 29 |
| 28 | 27 |
| 27 | 21 |
| 38 | 31 |
| 38 | 31 |
| 33 | 30 |
| 43 | 38 |
| 41 | 23 |
| 40 | 25 |
| 39 | 35 |
| 32 | 23 |
| 38 | 31 |
| 41 | 23 |
| 38 | 31 |
| 40 | 25 |
| 36 | 23 |
| 43 | 38 |
| 50 | 31 |
| 43 | 38 |
| 40 | 25 |
| 40 | 25 |
| 38 | 31 |
| 33 | 30 |
| 37 | 29 |
| 33 | 30 |
| 33 | 30 |
| 50 | 31 |
| 47 | 32 |
| 30 | 19 |
| 37 | 28 |
| 31 | 25 |
| 38 | 31 |
| 30 | 19 |