Бабин Д.Н., Петюшко А.А., Иванов И.Е.

Спецкурс «Введение в компьютерный интеллект. Машинное обучение»

**Метод наименьших квадратов.**

Рассмотрим задачу проведения прямой на плоскости, наименее уклоняющейся от заданных точек. Допустим, что точки результат измерений с погрешностью, а нам требуется восстановить прямую, считая погрешности минимальными. Пусть точки - это

 ,

а прямая задаётся уравнением



где  и  неизвестны. Ошибка в -той точке будет

.

Будем считать общей ошибкой сумму квадратов ошибок во всех точках и ее минимизировать.



Возьмем частные производные от  по  и , получим







.

Чтобы найти точку экстремума, приравняем частные производные к нулю. Получим систему уравнений для нахождения коэффициентов прямой наименее уклоняющейся от n точек.

Пусть требуется написать уравнение такой прямой для точек





Подставляем в формулу: =5, =3, =4, =3,n=3, получаем систему

,

решение которой . График прямой  и точки показаны на рисунке.

**Задача прогноза уровня инфляции.**

Пусть инфляция в некоторой стране имела следующие значения

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Год | 02 (2002) | 04 (2004) | 06 (2006) | 08 (2008) |
| Инфляция % | 15 | 10 | 12 | 11 |

Какой уровень инфляции можно прогнозировать в 2010 году?

**Решение**

Будем решать задачу методом проведения прямой, наименее уклоняющейся от 4-х точек

(2,15), (4,10), (6,12), (8,11)

=4+16+36+64=120, =2+4+6+8=20,

=30+40+72+88=230, =15+10+12+11=48,

n=4, получаем систему

, или 

решение которой

.

Подставив x=10 в уравнение прямой



получим прогноз процента инфляции в 2010 – **9.5%**

**Решение переопределенных систем линейных уравнений методом наименьших квадратов. Вывод формулы для n=2.**

Часто бывает, что составленная система уравнений такова, что число уравнений больше числа неизвестных, система несовместна, а сами коэффициенты ее найдены неточно. Требуется найти оптимальное приближенной решение такой системы. Сделаем это методом наименьших квадратов. Пусть система имеет вид

 или 

Пусть , .

Будем считать общей ошибкой сумму квадратов ошибок во всех уравнениях и ее минимизировать ее как функцию от 



Тогда 







Приравняв к нулю частные производные, получим систему:

,

которая может быть записана в матричном виде так:



.

Формула верна в общем случае матрицы mxn.

Таким образом, решение переопределенной системы



с матрицей  и вектором свободных членов 

будет совпадать с решением системы nxn вида:



**Задача о входящем и исходящем потоке машин.**

Некоторое агентство хочет анонимно собрать информацию о том, сколько времени проводят в клиенты в ресторане быстрого питания. Агенты могут посчитать автомобили, въезжающие и выезжающие со стоянки ресторана. Данные приведены в таблице

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Интервал времени | **12-13** | **13-14** | **14-15** | **15-16** | **16-17** |
| **Число приехавших машин** | **100** | **200** | **400** | **400** | **600** |
| **Число уехавших машин** | **0** | **0** | **200** | **275** | **400** |

Пусть x – это доля машин, которые уехали через час после приезда, а y – доля тех, кто приехал и уехал в тот же час.

Тогда 100 x – число машин приехавших в интервал с 12 до 13 часов и уехавших в интервал с 13 до 14 часов. 200 у – число машин приехавших в интервал с 13 до 14 часов и уехавших в интервал с 13 до 14 часов. Получаем

100 х + 200 у =0

200 x – число машин приехавших в интервал с 13 до 14 часов и уехавших в интервал с 14 до 15 часов. 400 у – число машин приехавших в интервал с 14 до 15 часов и уехавших в интервал с 14 до 15 часов. Получаем

200 х + 400 у =200

400 x – число машин приехавших в интервал с 14 до 15 часов и уехавших в интервал с 15 до 16 часов. 400 у – число машин приехавших в интервал с 15 до 16 часов и уехавших в интервал с 15 до 16 часов. Получаем

400 х + 400 у =275

400 x – число машин приехавших в интервал с 16 до 17 часов и уехавших в интервал с 16 до 17 часов. 400 у – число машин приехавших в интервал с 16 до 17 часов и уехавших в интервал с 16 до 17 часов. Получаем

400 х + 600 у =400

Сократив на 100 получим переопределенную систему линейных уравнений.

 или ,



Система 

имеет вид 

  



Доля тех, кто приехал, пообедал и уехал в тот же час – это 25%. А доля тех, кто просидел в ресторане около часа – 50%. Из этой информации можно делать заключить, что половина посетителей в ресторане не только обедает, но и общается. Если хозяин ресторана быстрого питания захочет увеличить пропускную способность своего заведения, ему надо создавать некомфортные условия для общения, такие как громкая музыка, яркий свет, высокие стулья и столы, и.т.д.