Intelligent Data Analysis Summer 2014
C. Braune

# Exercise Sheet 10

## Exercise 35 Logistic Regression

The following table shows the number of American intercontinental ballistic missiles (ICBMs) in the years from 1960 to 1969:

| year, $x$ | 1960 | 1961 | 1962 | 1963 | 1964 | 1965 | 1966 | 1967 | 1968 | 1969 |
|---|---|---|---|---|---|---|---|---|---|---|
| number, $y$ | 18 | 63 | 294 | 424 | 834 | 854 | 904 | 1054 | 1054 | 1054 |

Find a best fit curve for this data set using logistic regression ($Y = 1060$)! Draw the original data and sketch the curve $y = \frac{1060}{1+e^{a+bx}}$!

---

## Additional Exercise Exponential Regression

Radioactive substances decay according to the law $N(t) = N_0 e^{-\lambda t}$, where $t$ is the time, $\lambda$ a substance-specific decay parameter, $N_0$ the number of atoms of the substance at the beginning and $N(t)$ the number of remaining atoms at time point $t$. With the help of Geiger–Müller counter the following values $n$ were measured for the number of $\alpha$ particles that were emitted by a small amount of a radioactive substance up to different time points $t$:

| $t$ (in s) | 0 | 30 | 60 | 90 | 120 | 150 | 180 | 210 | 240 |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | 0 | 306 | 552 | 655 | 768 | 863 | 901 | 919 | 956 |

Each counted $\alpha$ particle indicates that one atom of the radioactive substance decayed. Determine the half-life of the radioactive substance! What element is this substance?

Procedure: Find a best fit curve $n = n_0(1 - e^{a+bt})$!
(Hint: You have to find a transformation that reduces the problem to the problem of finding a best fit line (regression line); $n_0 = 1000$.) Although the value for $a$ may differ from zero with this approach, $-b$ may be seen as an approximation of the decay parameter $\lambda$, from which the half-life can easily be determined. The half-life of a substance is the time after which only half of the originally present atoms remain.

---

**Exercise 36** Frequent Itemset Mining

Please use the Apriori algorithm for solving this exercise!

a) Find the frequent/maximal/closed item sets for the following transaction vector and $s_{min} = 3$:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1: | a | d | f | | | | | | |
| 2: | b | d | | | | | | | |
| 3: | b | c | | | | | | | |
| 4: | b | d | e | | | | | | |
| 5: | c | d | f | | | | | | |
| 6: | a | c | d | e | | | | | |
| 7: | b | c | d | | | | | | |
| 8: | a | b | d | | | | | | |
| 9: | b | c | e | g | | | | | |
| 10: | a | b | d | | | | | | |

b) Find an example of a transaction database for which the number of maximal item sets goes down if the minimum support is reduced; or explain in some other way why it is possible that the number of maximal item sets can also become smaller if the minimum support is reduced.