Homework 1

YueLi

15620161152260

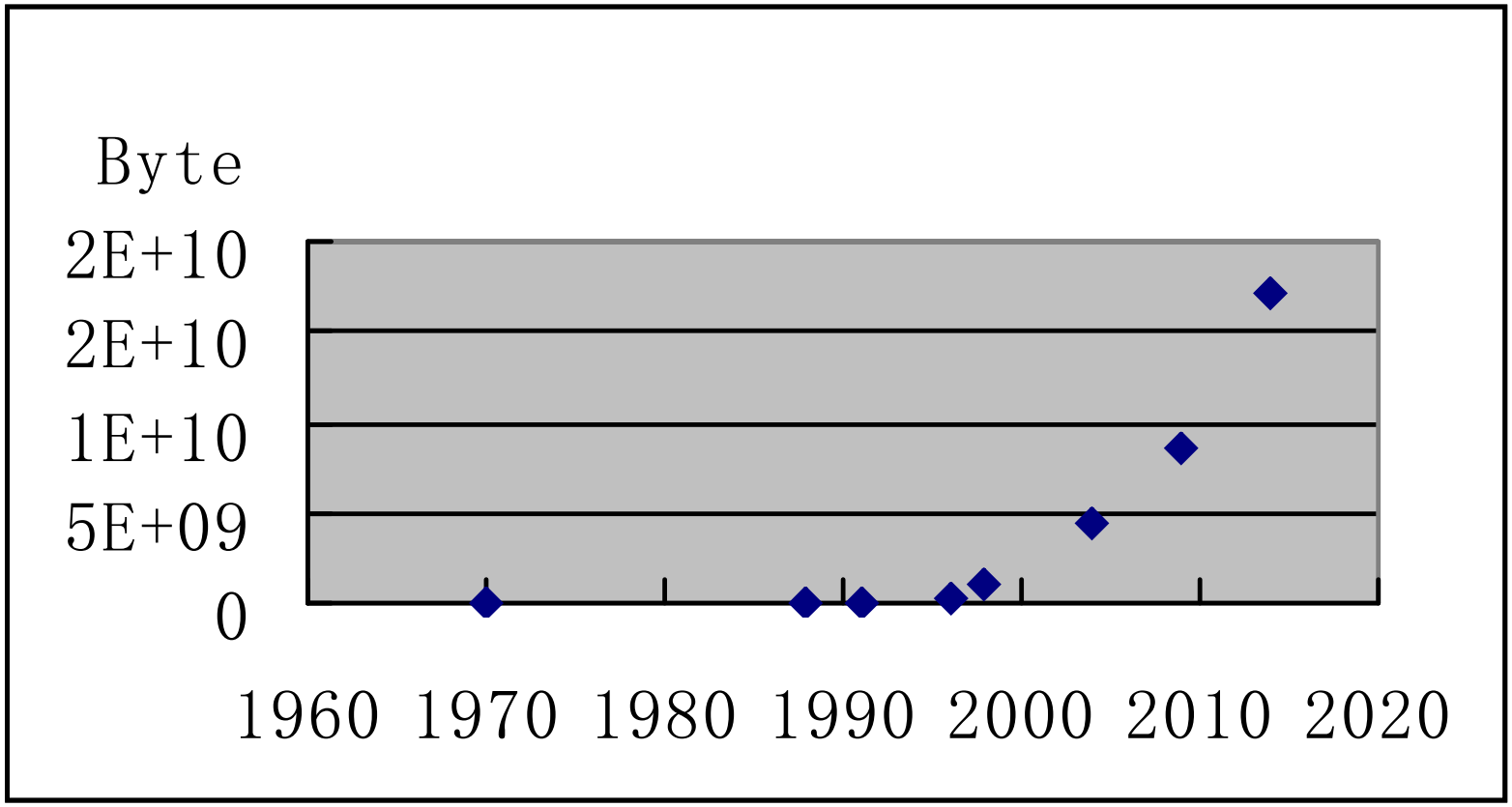
Q1

Date	Byte of PC memory
1970	262144
1988	2097152
1991	16777216
1996	2.68E+08
1998	1.07E+09
2004	4.29E+09
2009	8.59E+09
2014	1.72E+10

Q1

Date	Byte of PC memory
1970	262144
1988	2097152
1991	16777216
1996	2.68E+08
1998	1.07E+09
2004	4.29E+09
2009	8.59E+09
2014	1.72E+10

Q1



Q2 Logistic Regression

- Yue Li

1. Introduction

- A specific generalized linear model (GLM)
- A regression model where the dependent variable is categorical.
- Used when the dependent variable Y is binary
- (i.e., Y takes on only two values: "0" and "1", which represent outcomes such as pass/fail, win/lose, alive/dead or healthy/sick)
- Logistic regression allows you to predict the probability of a particular categorical response
- Examples of applications:
  - Identify risk factors associated with disease
  - Identify potentially important predictors in exploratory analyses of prospective clinical trials

2. Logistic Regression Model

Logistic Regression Model:

$$p = p(y = 1|x) = \frac{\exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_x)}{1 + \exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)}$$

P-probability

After logit transformation:

$$\log it(p) = \ln\left(\frac{p}{1-p}\right)$$

Get another form:

$$\log it(p) = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

which is linear in the predictors.

3. vs. other approaches

Logistic regression can be seen as a special case of GLM and thus analogous to linear regression.

The key differences between these two models:

- 1. The conditional distribution is a Bernoulli distribution rather than a Gaussian distribution, because the dependent variable is binary.
- 2. The predicted values are probabilities and are therefore restricted to (0,1) through the logistic distribution function .

THANKS