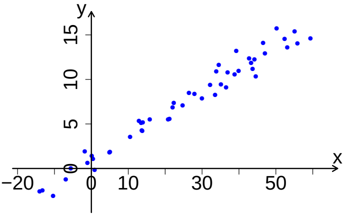
Lecture 1. From Linear Regression

Nan Ye

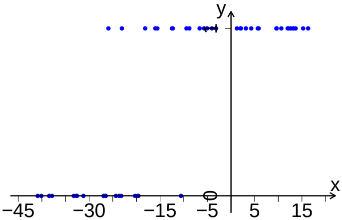
School of Mathematics and Physics
University of Queensland

Quiz

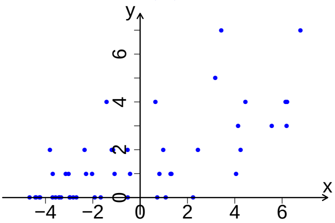
Q1. Which dataset is linear regression of y against x suitable for?



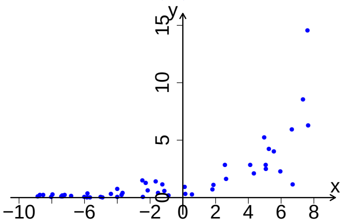
(a)



(b)



(c)



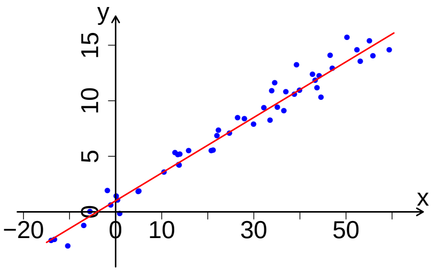
(d)

Q2. If there is a unique least squares regression line $y = \beta^\top \mathbf{x}$ on $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n) \in \mathbf{R}^d \times \mathbf{R}$, what is β?

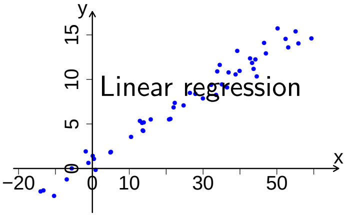
(a) $(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ (b) $(\mathbf{X}\mathbf{X}^\top)^{-1} \mathbf{X}\mathbf{y}$

(c) $\mathbf{X}^\top \mathbf{y}$ (d) $\mathbf{X}\mathbf{y}$

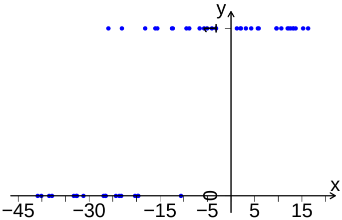
where $\mathbf{X}$ is the $n \times d$ design matrix with $\mathbf{x}_i$ as the i-th row, and $\mathbf{y} = (y_1, \ldots, y_n)^\top$.



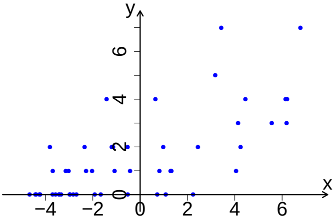
Q3. Suggest possible models for the data shown in the figures.



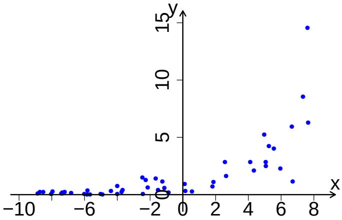
(a) Continuous



(b) Binary

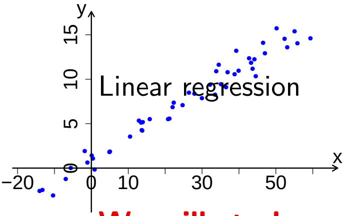


(c) Cardinal

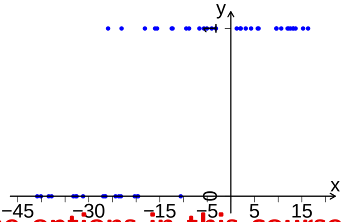


(d) Nonnegative continuous

Q3. Suggest possible models for the data shown in the figures.

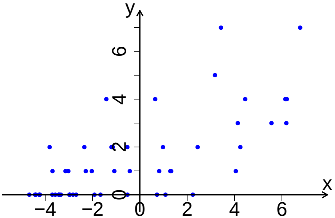


(a) Continuous

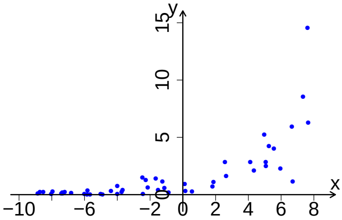


We will study some options in this course!

(b) Binary



(c) Cardinal



(d) Nonnegative continuous

Your Tasks

Assignment 4

14%	out 18 Sep, due 12pm 2 Oct

Assignment 5

14%	out 2 Oct, due 12pm 16 Oct

Consulting Project

	project description + data, out
2.5%	half-time check, due 6pm 1 Oct
7.5%	seminar, during a lecture in the week of 22 Oct
20%	report, due 6pm on 26 Oct

There are bonus questions in lectures and assignments.

Our Problem

Regression

Course Objective

- Understand the general theory of generalized linear models

  model structure, parameter estimation, asymptotic normality, prediction
- Be able to *recognize* and *apply* generalized linear models and extensions for regression on different types of data
- Be able to determine the goodness of fit and the prediction quality of a model

Put it simply, to be able to do regression using generalized linear models and extensions...

Course Overview

Generalized linear models (GLMs)

- Building blocks

  systematic and random components, exponential familes

- Prediction and parameter estimation
- Specific models for different types of data

  continuous response, binary response, count response...

- Modelling process and model diagnostics

Extensions of GLMs

- Quasi-likelihood models
- Nonparametric models
- Mixed models and marginal models

Time series

This Lecture

- Revisit basics of OLS
- Systematic and random components of OLS
- Extensions of OLS to other types of data
- A glimpse on generalized linear models

Revisiting OLS

The objective function

Ordinary least squares (OLS) finds a hyperplane minimizing the sum of squared errors (SSE)

$$\beta_n = \arg\min_{\beta \in \mathbf{R}^d} \sum_{i=1}^{n} (\mathbf{x}_i^\top \beta - y_i)^2,$$

where each $\mathbf{x}_i \in \mathbf{R}^d$ and each $y_i \in \mathbf{R}$.

Terminology

$\mathbf{x}$*: input, independent variables, covariate vector, observation, predictors, explanatory variables, features.*

y*: output, dependent variable, response.*

Solution

The solution to OLS is

$$\beta_n = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y},$$

where $\mathbf{X}$ is the $n \times d$ design matrix with $\mathbf{x}_i$ as the i-th row, and $\mathbf{y} = (y_1, \ldots, y_n)^\top$.

The formula holds when $\mathbf{X}^\top \mathbf{X}$ is non-singular. When $\mathbf{X}^\top \mathbf{X}$ is singular, there are infinitely many possible values for β_n. They can be obtained by solving the linear systems $(\mathbf{X}^\top \mathbf{X})\beta = \mathbf{X}^\top \mathbf{y}$.

Justification as MLE

- Assumption: $y_i \mid \mathbf{x}_i \overset{ind}{\sim} N(\mathbf{x}_i^\top \beta, \sigma^2)$.
- Derivation: the log-likelihood of β is given by

$$
\begin{aligned}
&\ln p(y_1, \dots, y_n \mid \mathbf{x}_1, \dots, \mathbf{x}_n, \beta) \\
&= \sum_i \ln p(y_i \mid \mathbf{x}_i, \beta) \\
&= \sum_i \ln \left(\frac{1}{\sqrt{2\pi\sigma}} \exp(-(y_i - \mathbf{x}^\top \beta)^2 / 2\sigma^2) \right) \\
&= \text{const.} - \frac{1}{\sigma^2} \sum_i (y_i - \mathbf{x}_i^\top \beta)^2.
\end{aligned}
$$

Thus minimizing the SSE is the same as maximizing the log-likelihood, i.e. maximum likelihood estimation (MLE).

An Alternative View

- OLS has two orthogonal components

$$\text{(systematic)} \quad \mathbb{E}(Y \mid \mathbf{x}) = \beta^\top \mathbf{x}.$$

$$\text{(random)} \quad Y \mid \mathbf{x} \text{ is normally distributed with variance } \sigma^2.$$

- This has two key features
  - Expected value of Y given $\mathbf{x}$ is a function of $\beta^\top \mathbf{x}$.
  - Parameters of the conditional distribution of Y given $\mathbf{x}$ can be determined from $\mathbb{E}(Y \mid \mathbf{x})$.
- This defines a conditional distribution $p(y \mid \mathbf{x}, \beta)$, with parameters estimated using MLE.

Generalization

(systematic) $\mathbb{E}(Y \mid \mathbf{x}) = g(\beta^\top \mathbf{x})$.

(random) $Y \mid \mathbf{x}$ is normally/Poisson/Bernoulli/... distributed.

Example 1. Logistic regression for binary response

- When Y takes value 0 or 1, we can use the logistic function to squash $\mathbf{x}^\top \beta$ to $[0, 1]$, and use the Bernoulli distribution to model $Y \mid \mathbf{x}$, as follows.

$$\text{(systematic)} \quad \mathbb{E}(Y \mid \mathbf{x}) = \textit{logistic}(\beta^\top \mathbf{x}) = \frac{1}{1 + e^{-\beta^\top \mathbf{x}}}.$$
$$\text{(random)} \quad Y \mid \mathbf{x} \text{ is Bernoulli distributed.}$$

- Or more compactly,

$$Y \mid \mathbf{x} \sim B\left(\frac{1}{1 + e^{-\beta^\top \mathbf{x}}}\right),$$

where $B(p)$ is the Bernoulli distribution with parameter p.

Example 2. Poisson regression for count response

- When Y is a count, we can use exponentiation to map $\beta^\top \mathbf{x}$ to a non-negative value, and use the Poisson distribution to model $Y \mid \mathbf{x}$, as follows.

$$\begin{aligned} \text{(systematic)} \quad & \mathbb{E}(Y \mid \mathbf{x}) = \exp(\beta^\top \mathbf{x}). \\ \text{(random)} \quad & Y \mid \mathbf{x} \text{ is Poisson distributed.} \end{aligned}$$

- Or more compactly,

$$Y \mid \mathbf{x} \sim Po\left(\exp(\beta^\top \mathbf{x})\right),$$

where $Po(\lambda)$ is a Poisson distribution with parameter λ.

Example 3. Gamma regression for non-negative response

- When Y is a non-negative continuous random variable, we can choose the systematic and random components as follows.

$$\begin{aligned} \text{(systematic)} \quad & \mathbb{E}(Y \mid \mathbf{x}) = \exp(\beta^\top \mathbf{x}) \\ \text{(random)} \quad & Y \mid \mathbf{x} \text{ is Gamma distributed.} \end{aligned}$$

- We further assume the variance of the Gamma distribution is μ^2/ν (ν treated as known), thus

$$Y \mid \mathbf{x} \sim \Gamma(\mu = \exp(\beta^\top \mathbf{x}), \text{var} = \mu^2/\nu),$$

where $\Gamma(\mu = a, \text{var} = b)$ denotes a Gamma distribution with mean a and variance b.

Generalized Linear Models

- A GLM has the following structure

$$\begin{aligned} \text{(systematic)} \quad & \mathbb{E}(Y \mid \mathbf{x}) = h(\beta^\top \mathbf{x}). \\ \text{(random)} \quad & Y \mid \mathbf{x} \text{ follows an exponential family distribution.} \end{aligned}$$

- This is usually separated into three components
  - The linear predictor $\beta^\top \mathbf{x}$.
  - The response function h.
    People often specify the link function $g = h^{-1}$ instead.
  - The exponential family for the conditional distribution of Y given $\mathbf{x}$.

Remarks on the exponential family

- It is common!

  normal, Bernoulli, Poisson, and Gamma distributions are exponential families.

- It gives a well-defined model.

  its parameters are determined by the mean $\mu = \mathbb{E}(Y \mid \mathbf{x})$.

- It leads to a unified treatment of many different models.

  linear regression, logistic regression, ...

We will take a close look at these in the next few lectures.

What You Need to Know

- The approach of regression by separately specifying systematic and random components.
- Example applications of the approach
  - *Linear regression, logistic regression, Poisson regression, Gamma regression*
- The components of generalized linear models