Regression

Wenda Zhou

June 6, 2017

Regression

Relate the parameter of interest to covariates.

Example

Clinical trial Suppose that we wish to estimate the effectiveness of a drug to control blood pressure. We may model that as a Normal with mean μ, and variance σ^2.
However, we may believe that μ depends on covariates, such as age, sex, weight, etc.

Regression

Idea: write μ as a function of the covariates (such as age, sex, weight, etc.):

$$\mu = f(\text{age}, \text{weight}, \dots) \tag{1}$$

Regression

Idea: write μ as a function of the covariates (such as age, sex, weight, etc.):

$$\mu = f(\text{age}, \text{weight}, \dots) \tag{1}$$

What should that function be?
Estimate the function from data.

Linear regression

Consider a special case: g is a linear function.

$$f(\text{age}, \text{weight}) = \alpha + \beta_{\text{age}} \times \text{age} + \beta_{\text{weight}} \times \text{weight} \qquad (2)$$

parameters are α and β

Linear regression

Consider a special case: g is a linear function.

$$f(\text{age}, \text{weight}) = \alpha + \beta_{\text{age}} \times \text{age} + \beta_{\text{weight}} \times \text{weight} \qquad (2)$$

parameters are α and β
This is linear regression.

Linear regression

Linear regression is among the most used statistical models.

Strong points

- Easy to interpret
- Flexible
- Does not require too much data

Weak points

- Can be restrictive
- Sometimes too simple

Ordinary least squares

Special interpretation for the case of a normal model.
Likelihood for i.i.d. normal $y_1, \ldots, y_n$ is given by:

$$\ell(\mu) = -\sum_{i=1}^{n} \frac{1}{2\sigma^2}(y_i - \mu)^2 - \frac{n}{2}\log(\pi\sigma^2) \tag{3}$$

$(y_i - \mu)^2$ is squared distance from point to line.

Simple linear regression

Suppose have only a single covariate x.
Observations (y_i, x_i).
Compute the m.le. estimator of α and β where:

$$\mu = \alpha + \beta x \tag{4}$$

Simple linear regression

Suppose have only a single covariate x.
Observations (y_i, x_i).
Compute the m.le. estimator of α and β where:

$$\mu = \alpha + \beta x \tag{4}$$

$$\begin{cases} \hat{\beta} = \frac{\mathrm{Cov}(x,y)}{\mathrm{Var}\, x} \\ \hat{\alpha} = \bar{y} - \bar{x}\hat{\beta} \end{cases} \tag{5}$$

Residuals

Define the residual r_i to be:

$$r_i = y_i - (\alpha + \beta x_i) \tag{6}$$

Can help us characterise how adequate the fit is.

Residuals

Define the residual r_i to be:

$$r_i = y_i - (\alpha + \beta x_i) \tag{6}$$

Can help us characterise how adequate the fit is.
Indeed, $\hat{\alpha}$ and $\hat{\beta}$ minimize

$$SS_{\text{res}} = r_i^2 \tag{7}$$

Residuals

Define the coefficient of determination:

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} \tag{8}$$

where $SS_{\text{tot}} = \sum_{i=1}^{n} (y_i - \bar{y})^2$.

Generalized linear models

Suppose that we wish to model a clinical trial where we measure the treament failure or success.
The outcome is no longer normal, but Bernoulli.

Generalized linear models

Suppose that we wish to model a clinical trial where we measure the treament failure or success.
The outcome is no longer normal, but Bernoulli.
Parameter of interest is p, probability of success.

Generalized linear models

As in the other linear models, may try to put:

$$p = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots \tag{9}$$

Generalized linear models

As in the other linear models, may try to put:

$$p = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots \tag{9}$$

However, we must also ensure $0 \leq p \leq 1$.
Idea: use a link function.

Generalized linear models

Model a function of p, that is:

$$g(p) = \alpha + \beta x \tag{10}$$

or equivalently,

$$p = g^{-1}(\alpha + \beta x) \tag{11}$$

Generalized linear models

Model a function of p, that is:

$$g(p) = \alpha + \beta x \tag{10}$$

or equivalently,

$$p = g^{-1}(\alpha + \beta x) \tag{11}$$

g is called the link function.

Link functions

The most common choice for a Bernoulli model is the logit link, which gives logistic regression.

$$g(p) = \text{logit}\, p = \log\left(\frac{p}{1-p}\right) \tag{12}$$

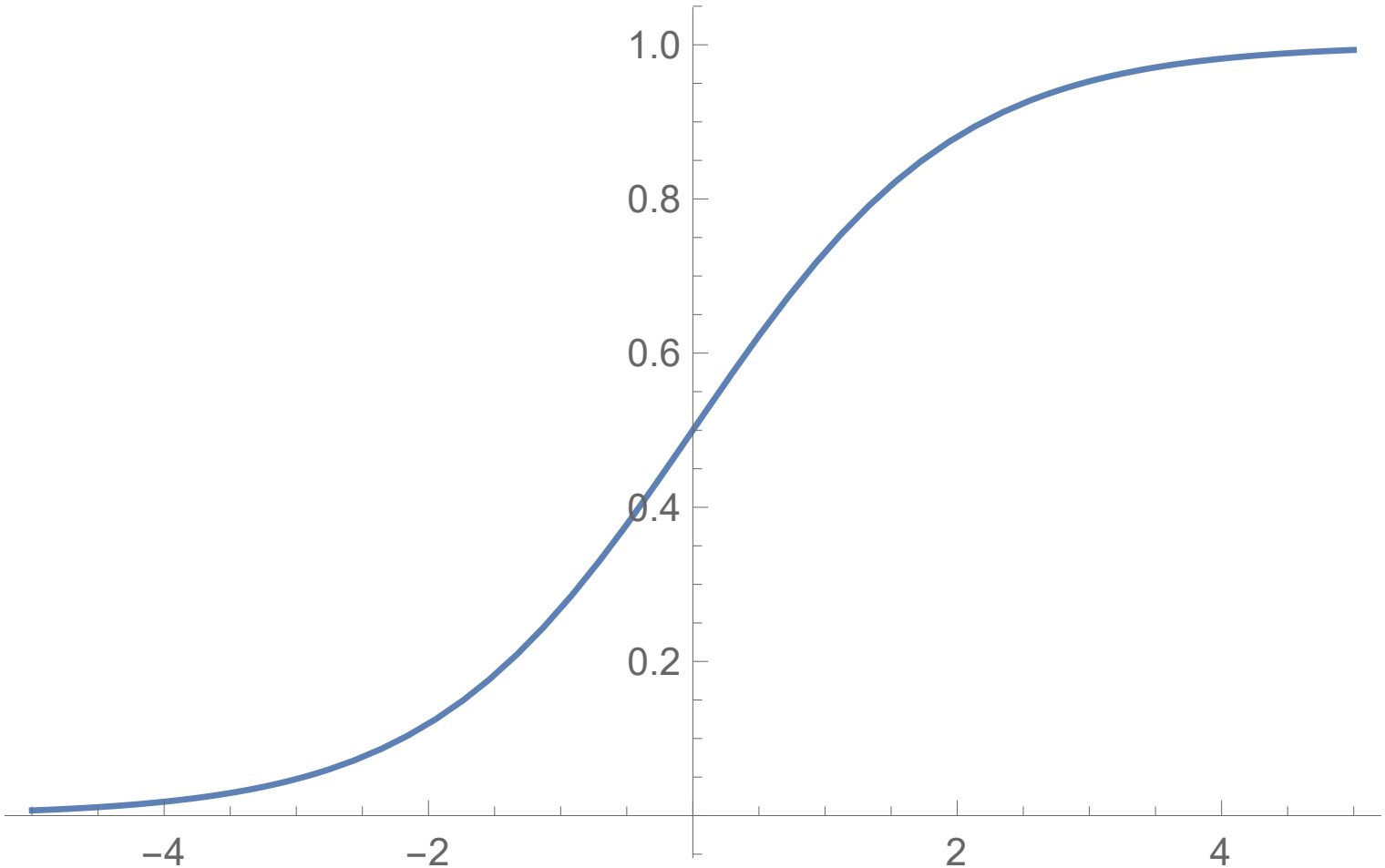
Its inverse is the sigmoid:

$$g^{-1}(x) = \frac{e^x}{1+e^x} \tag{13}$$

Link functions

Logit

The inverse-logit is increasing, and has value 0 at $-\infty$ and value 1 and $+\infty$.



Link functions

The most commonly used distributions that require a link are the Bernoulli and Poisson distribution.

Bernoulli

Parameter p, $0 \leq p \leq 1$.
Usual link is logit:

$$g(p) = \log\left(\frac{p}{1-p}\right) \tag{14}$$

Poisson

Parameter $\lambda \geq 0$.
Usual link is logarithmic:

$$g(\lambda) = \log \lambda \tag{15}$$

Non-linear regression

Sometimes may feel that linear assumption is too restrictive.
Numerous models attempt to generalise the linear assumption.

Generalized additive models

The generalized additive models are one attempt, and model

$$g(\theta) = \alpha + f_1(x_1) + f_2(x_2) + \cdots \tag{16}$$

where $f_1, \ldots$ are estimated from the data.

Generalized additive models

The generalized additive models are one attempt, and model

$$g(\theta) = \alpha + f_1(x_1) + f_2(x_2) + \cdots \tag{16}$$

where $f_1, \ldots$ are estimated from the data.
Similar but more complex maximum likelihood strategy.