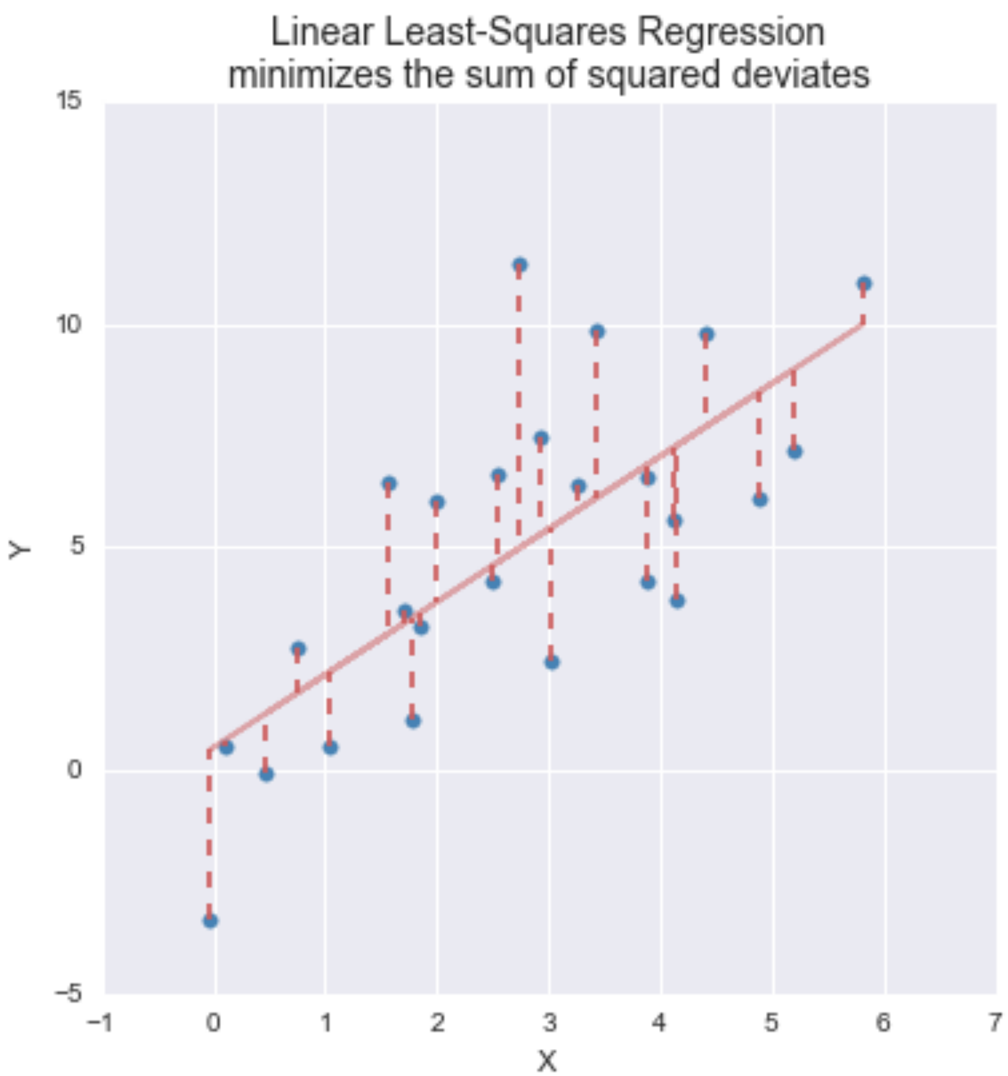
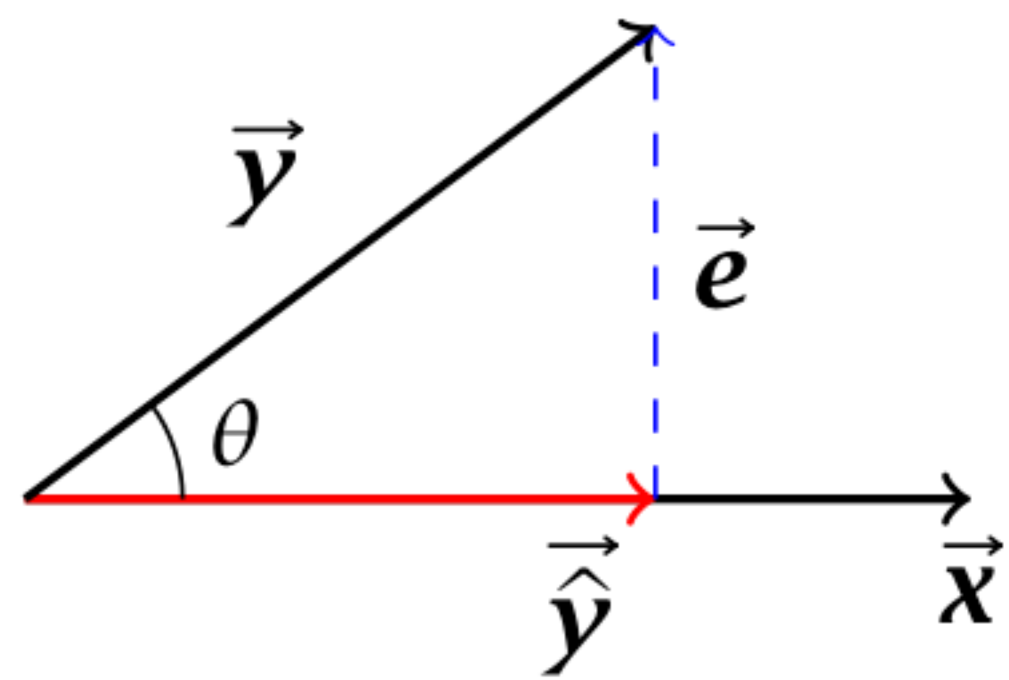
Additional Regression Models

Paul M. Magwene

Review: Bivariate Regression

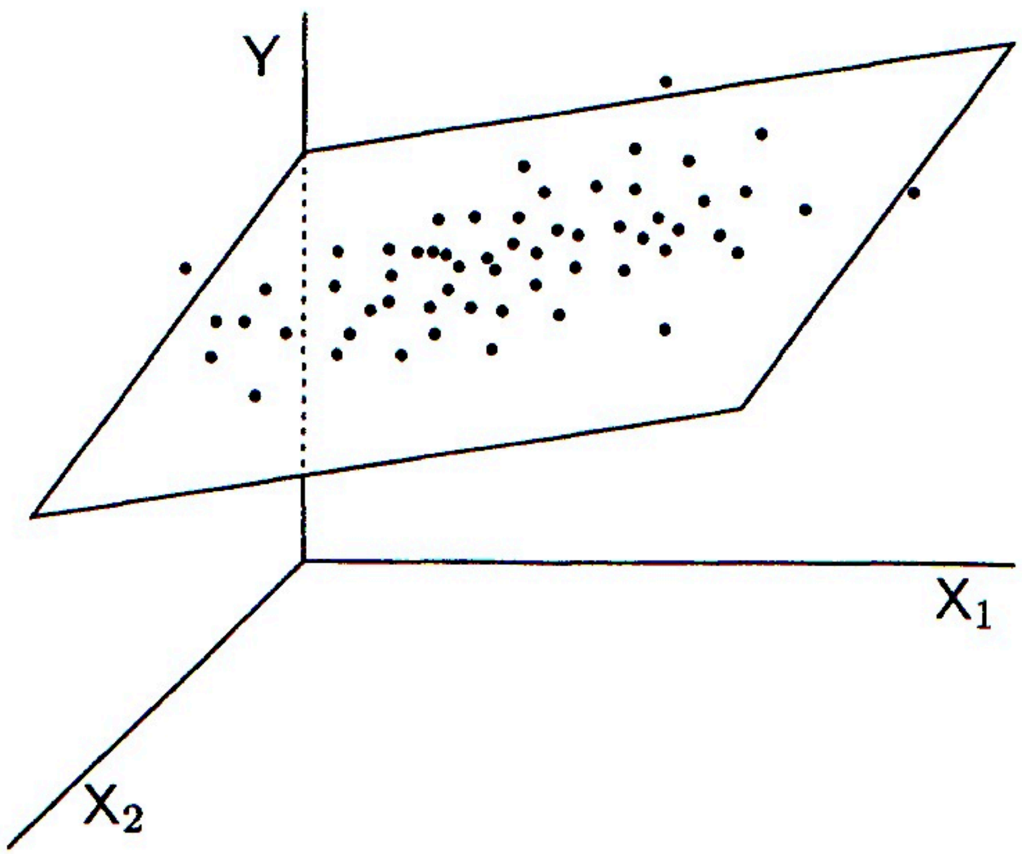


Variable space representation

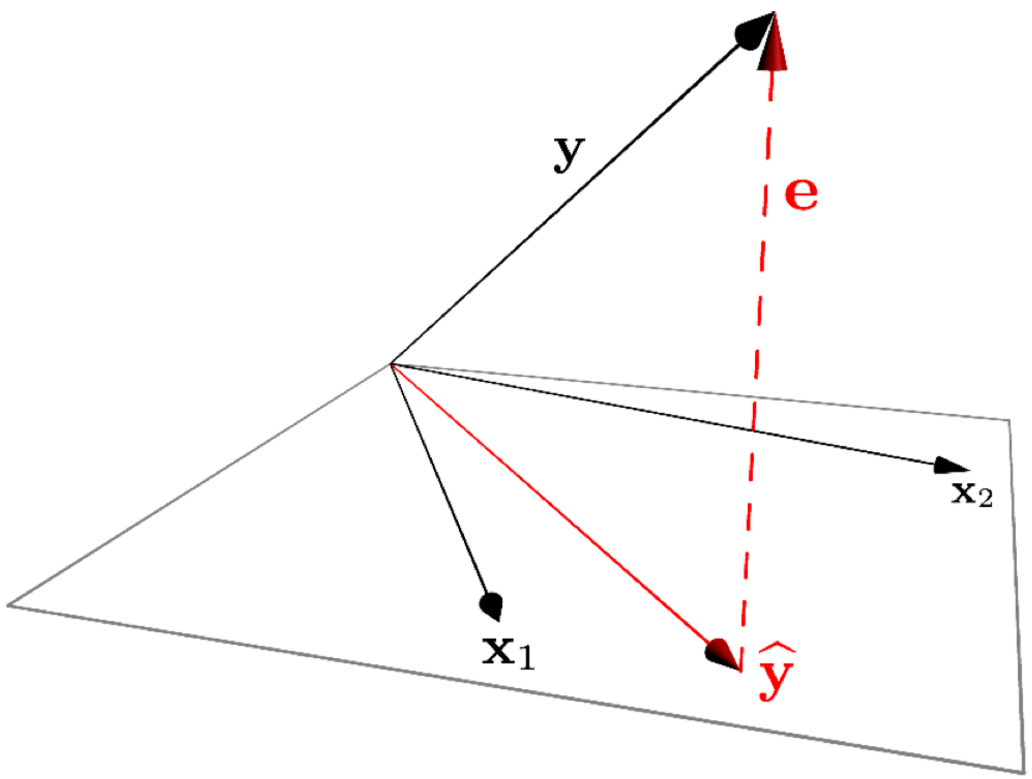


Subject space representation

Review: Multiple Regression

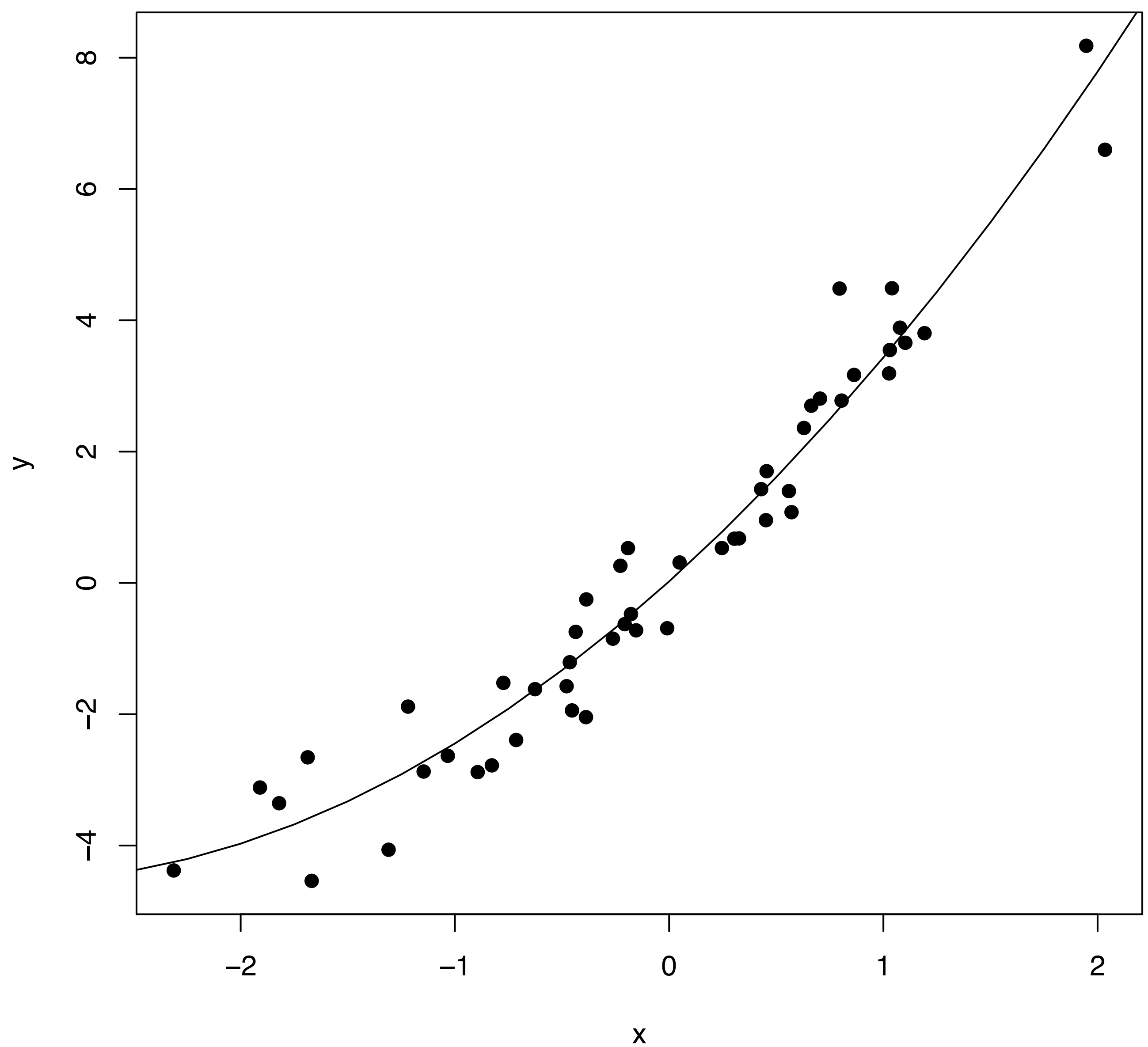


Variable space representation



Subject space representation

Curvilinear (Polynomial) Regression



Curvilinear Regression

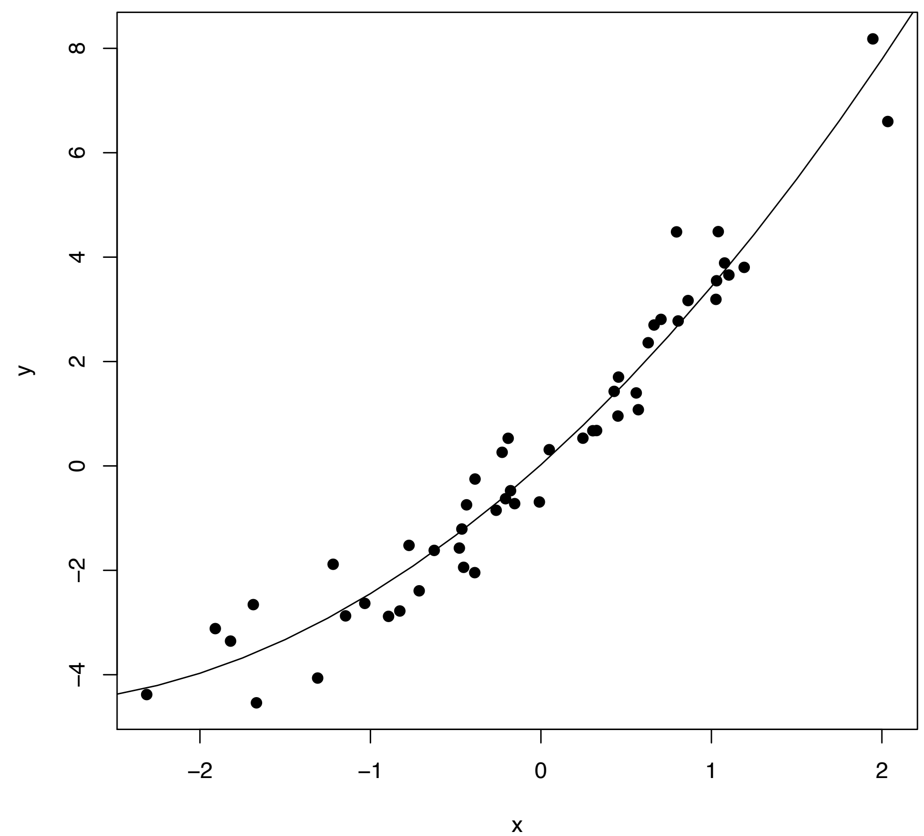
Curvilinear regression using **polynomial models** is simply multiple regression with the x_i replace by powers of x.

$$\hat{y} = b_1 \mathsf{x} + b_2 \mathsf{x}^2 + \cdots + b_p \mathsf{x}^n$$

Note:

- this is still a *linear* regression (linear in the coefficients)
- best applied when a specific hypothesis justifies there use
- generally not higher than quadratic or cubic

Curvilinear regression: synthetic example

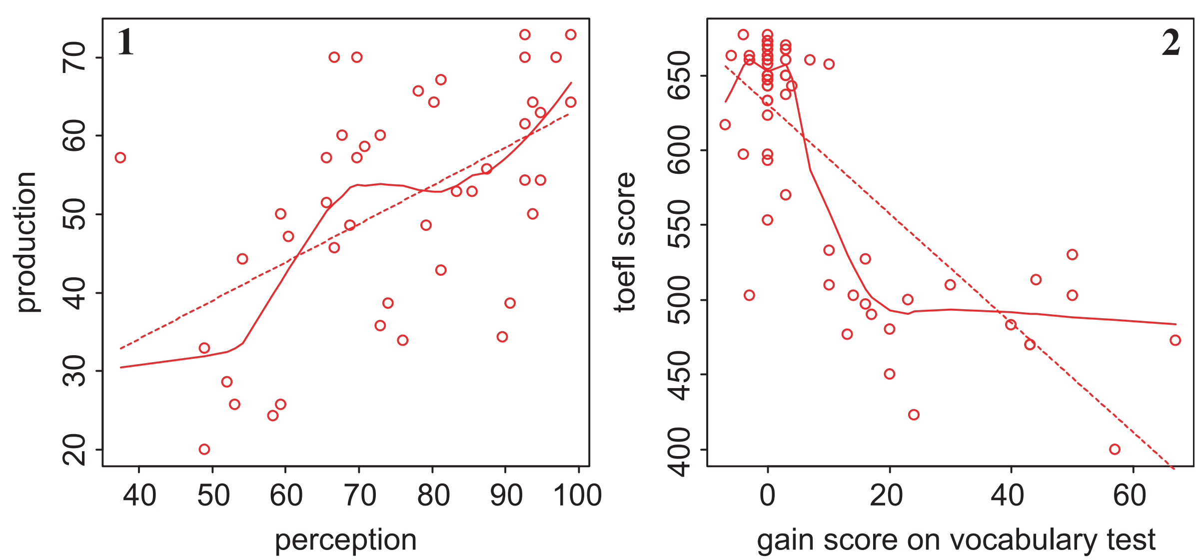


$$y = 3x + 0.5x^2 + e$$

```
lm(formula = y ~ x + I(x^2))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.02229    0.11651   0.191    0.849
x            2.94001    0.09693  30.331  < 2e-16 ***
I(x^2)       0.47146    0.07685   6.135 1.68e-07 ***
```

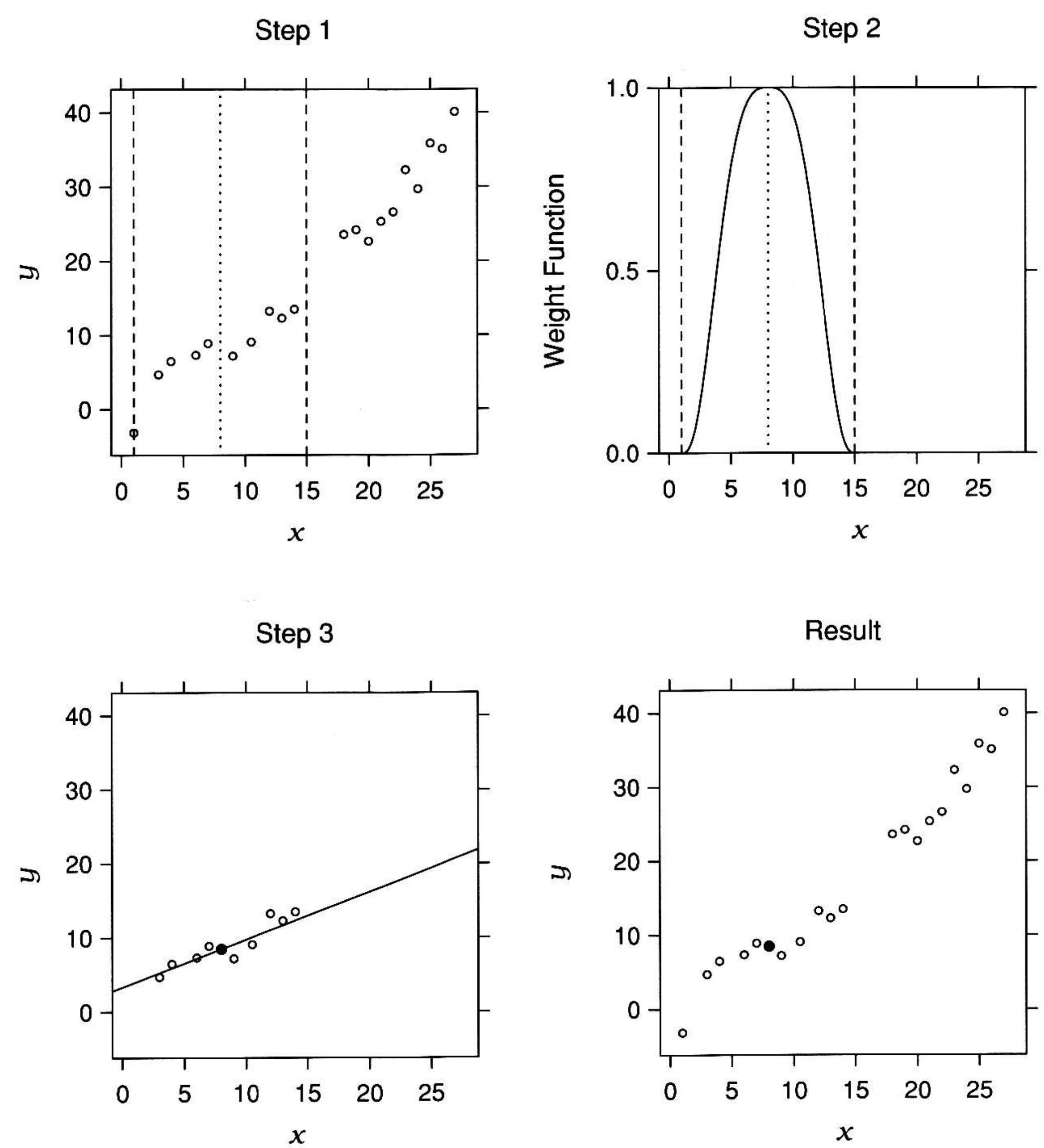
LOESS regression

- A type of non-parametric regression
- Basic idea – fit a curve (or surface) to a set of data by fitting a large number of *local regressions*.
- Cleveland, W.S. (1979). “Robust Locally Weighted Regression and Smoothing Scatterplots". Journal of the American Statistical Association 74 (368): 829-836. doi:10.2307/2286407.



LOESS, graphical overview I

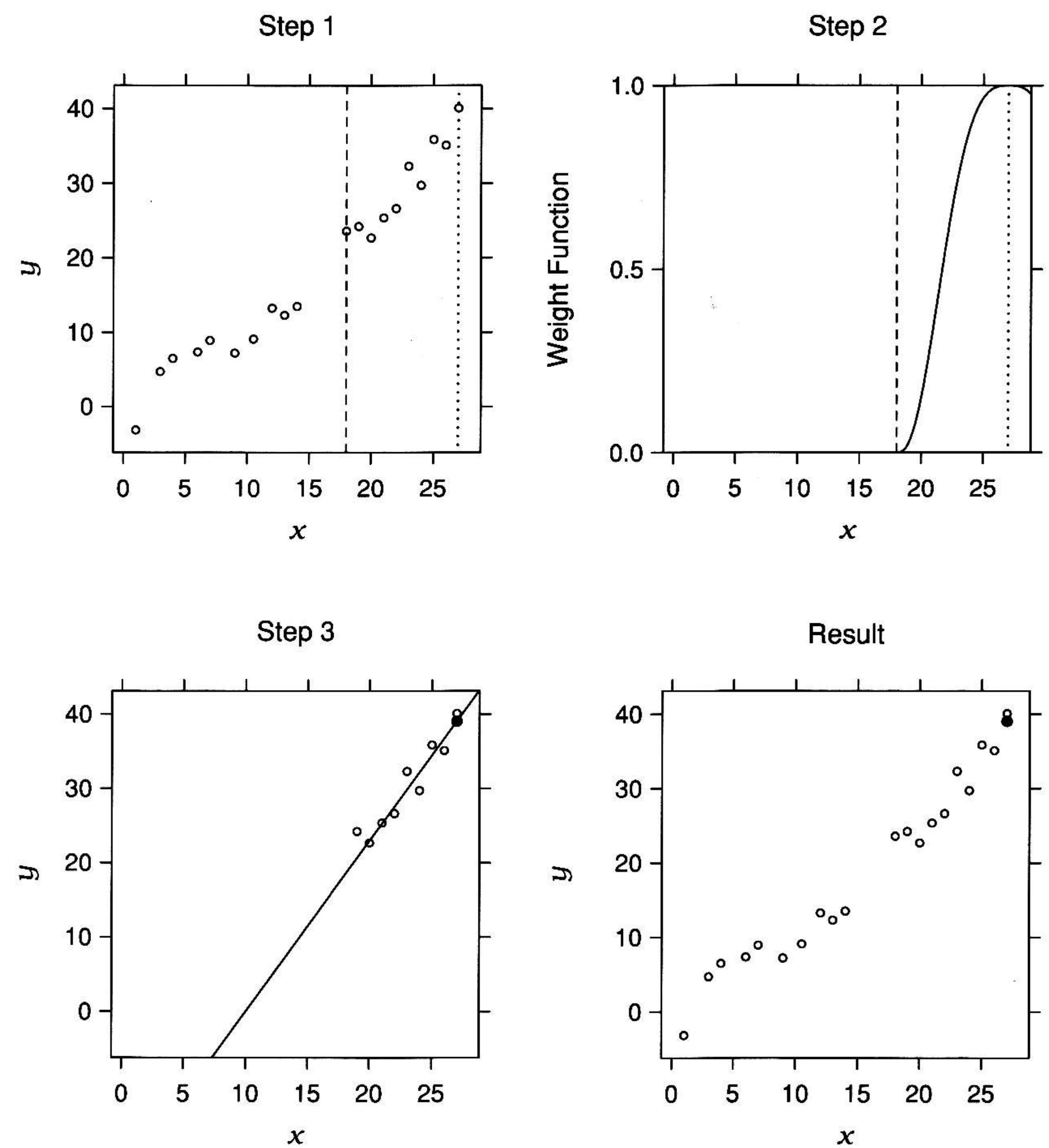
from Cleveland (1993)



3.49 HOW LOESS WORKS. The graphs show how the initial fit at x = 8 is computed. (Top left) α, which is 0.5, is multiplied by 20, the number of points, which gives 10. A vertical strip is defined around x = 8 so that one boundary is at the 10th nearest neighbor. (Top right) Weights are defined for the points using the weight function. (Bottom left) A line is fitted using weighted least-squares. The value of the line at x = 8 is the initial loess fit at x = 8. (Bottom right) The result is one point of the initial loess curve, shown by the filled circle.

LOESS, graphical overview II

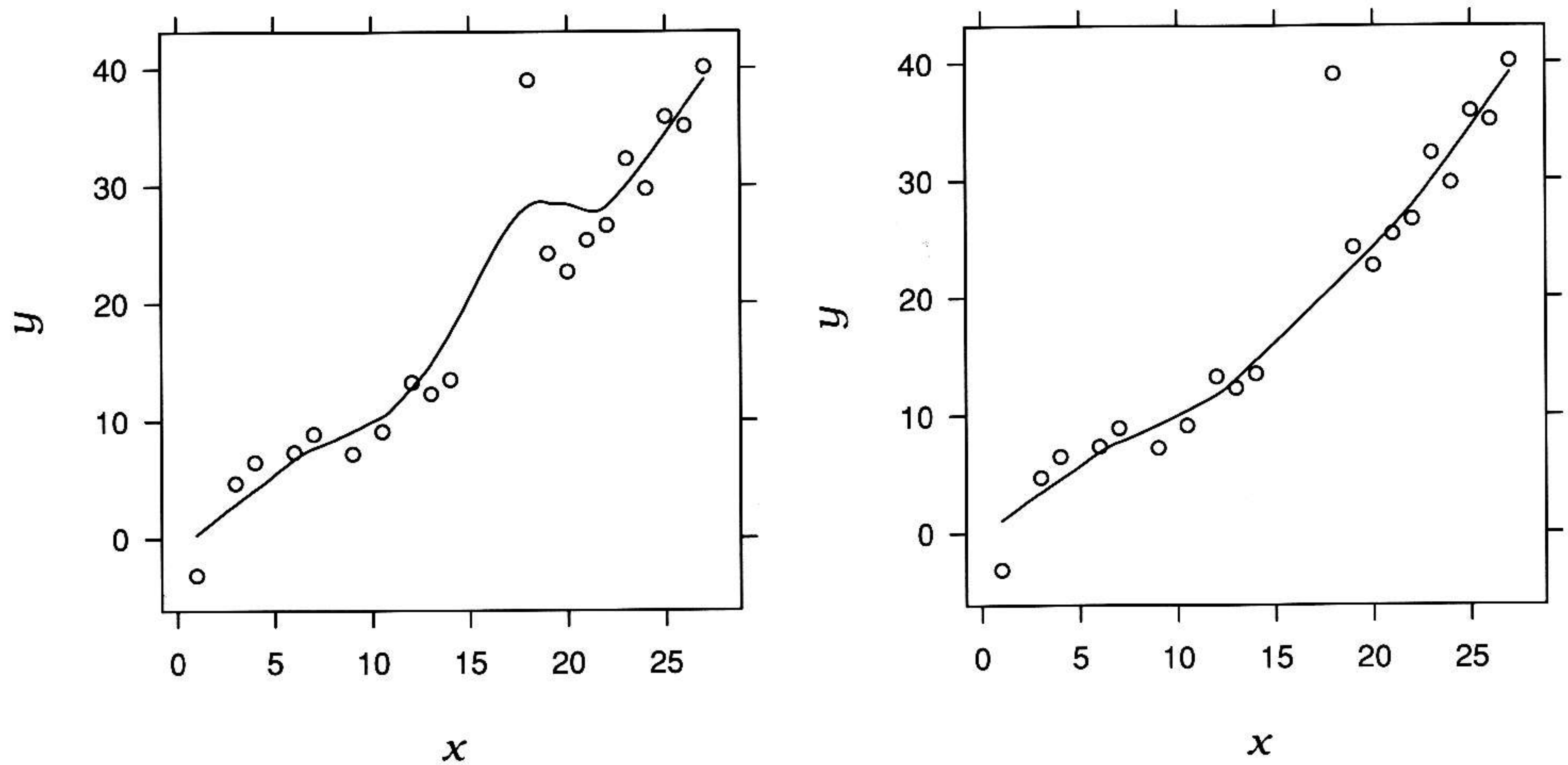
from Cleveland (1993)



3.50 HOW LOESS WORKS. The computation of the initial loess fit value at $x = 27$ is illustrated.

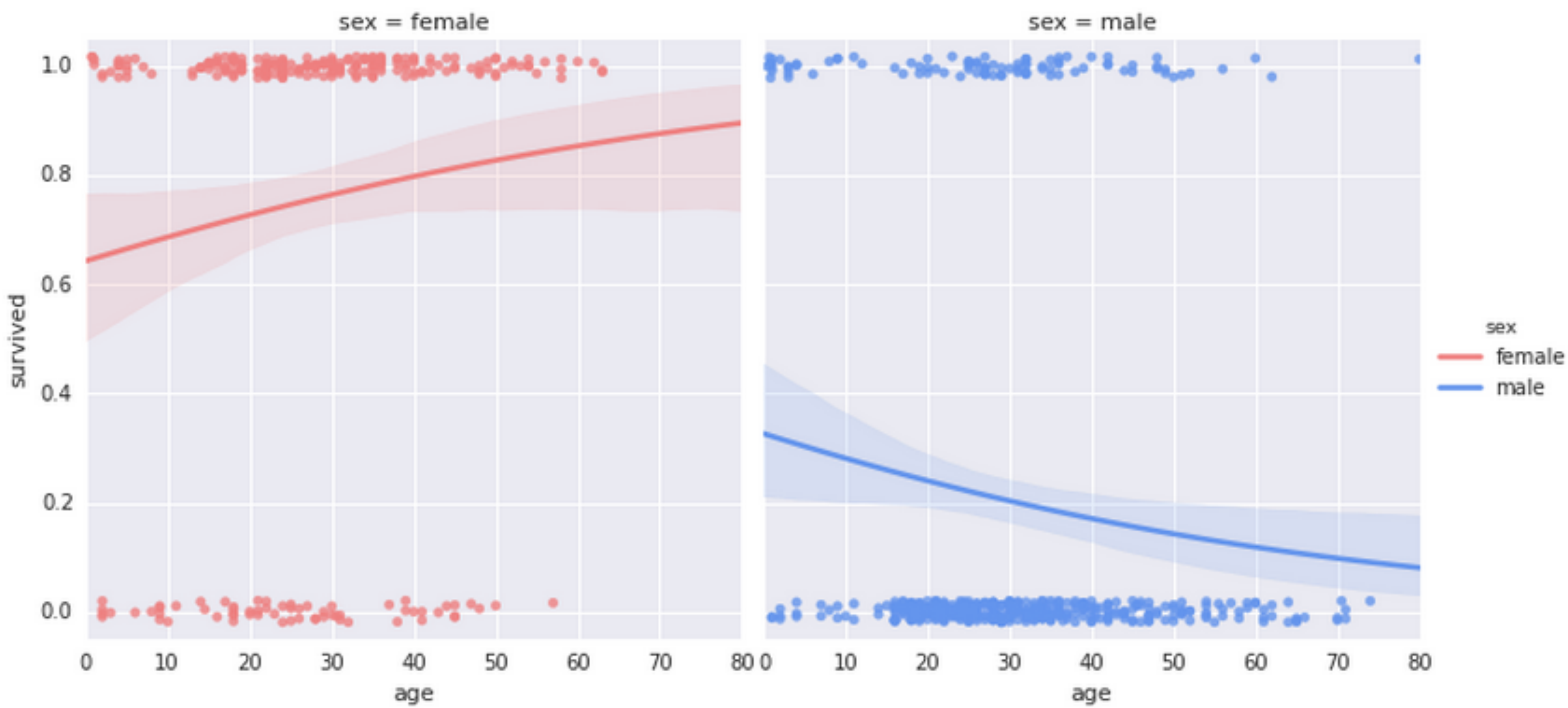
LOESS, graphical overview III

from Cleveland (1993)



3.51 HOW LOESS WORKS. Loess employs robustness iterations that prevent outliers from distorting the fit. (Left panel) The open circles are the points of the graph; there is one outlier between x = 15 and x = 20. The initial loess curve has been distorted in the neighborhood of the outlier. (Right panel) The graphed curve is the fit after four robustness iterations. Now the fit follows the general pattern of the data.

Logistic regression



How did the probability of survival vary with age for passengers on the Titanic?

Logistic regression

Logistic regression is used when the dependent variable is discrete (often binary). The explanatory variables may be either continuous or discrete.

Examples:

- whether a gene is turned off (=0) or on (=1) as a function of levels of various proteins
- whether an individual is healthy (=0) or diseased (=1) as a function of various risk factors.
- whether an individual animal died (=0) or survived (=1) some selective event as a function of one or more moprhological traits.

Logistic regression: mathematical formulation

Model the binary responses as:

$$P(Y = 1|X_1, \ldots, X_p) = f(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)$$

So we're modeling the probability of the states as a function of a linear combination of the predictor variables.

For logistic regression, we use the logistic function for f:

$$f(z) = \frac{1}{1 + e^{-z}}$$

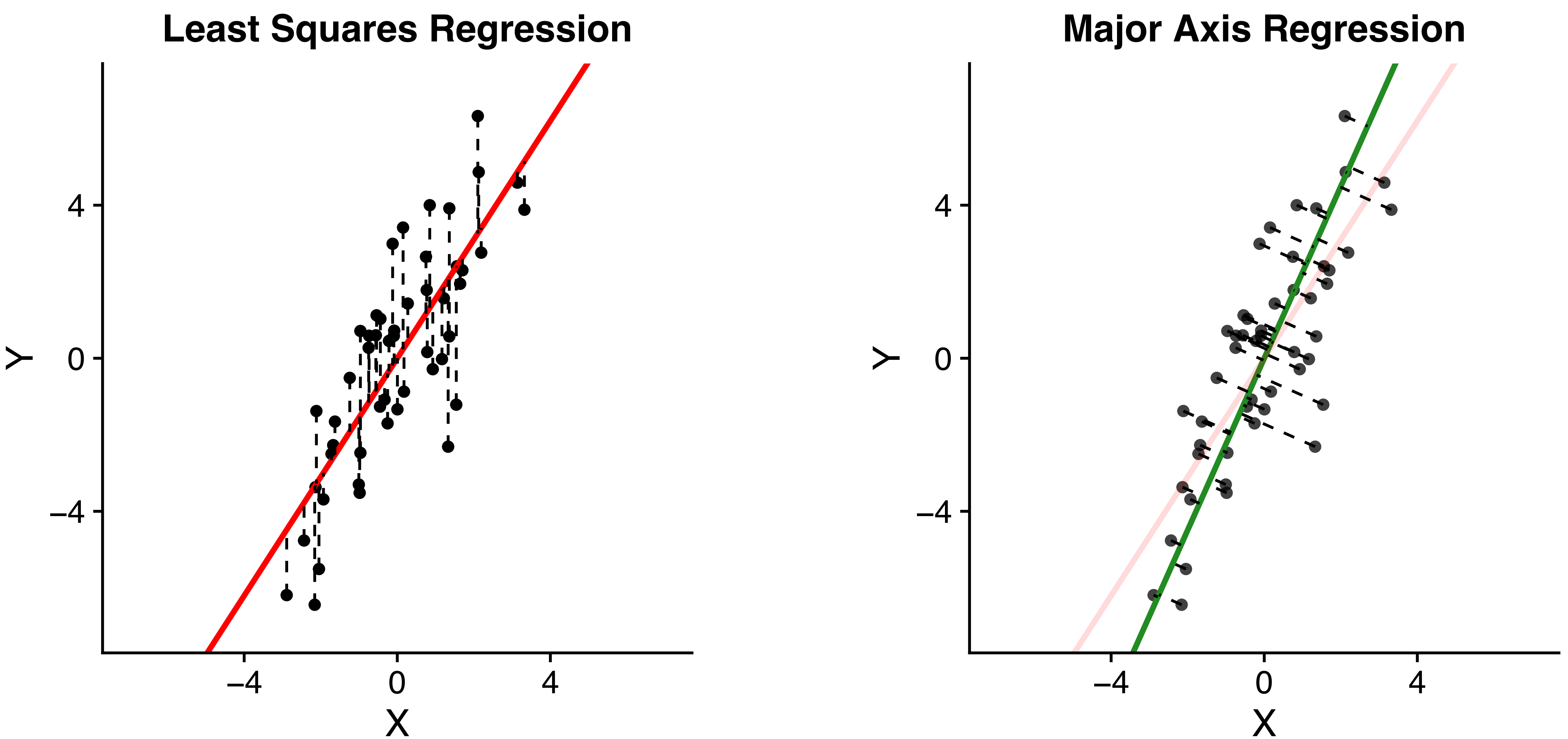
Substituting our linear combination of the predictor variables into the logistic function, for the bivariate case we get:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

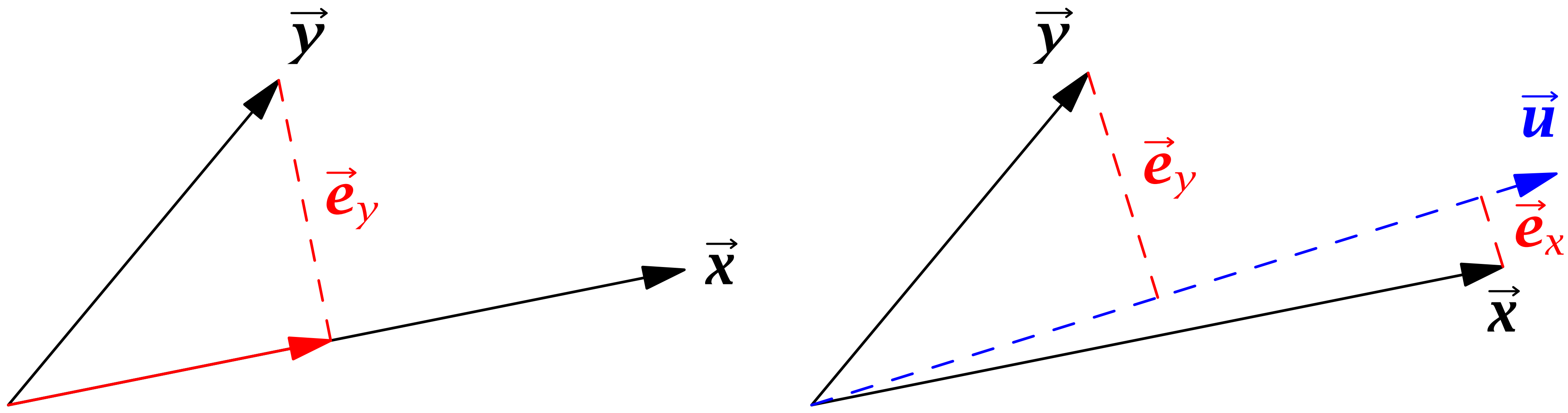
Notes on logistic regression

- The regression is no longer linear
- Estimating the β in logistic regression is done via maximum likelihood estimation (MLE)
- Information-theoretic metrics of model fit rather than F-statistics

What if we changed the optimality criterion for linear regression?



Vector geometry of least squares regression and major axis regression



(a) OLS

(b) Major Axis Regression

Figure: Vector geometry of ordinary least-squares and major axis regression.