

2. Linear functions

Outline

Linear and affine functions

Taylor approximation

Regression model

Superposition and linear functions

- ▶ $f : \mathbf{R}^n \rightarrow \mathbf{R}$ means f is a function mapping n -vectors to numbers
- ▶ f satisfies the *superposition property* if

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$$

holds for all numbers α, β , and all n -vectors x, y

- ▶ be sure to parse this very carefully!
- ▶ a function that satisfies superposition is called *linear*

The inner product function

- ▶ with a an n -vector, the function

$$f(x) = a^T x = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$$

is the *inner product function*

- ▶ $f(x)$ is a weighted sum of the entries of x
- ▶ the inner product function is linear:

$$\begin{aligned} f(\alpha x + \beta y) &= a^T (\alpha x + \beta y) \\ &= a^T (\alpha x) + a^T (\beta y) \\ &= \alpha (a^T x) + \beta (a^T y) \\ &= \alpha f(x) + \beta f(y) \end{aligned}$$

...and all linear functions are inner products

- ▶ suppose $f : \mathbf{R}^n \rightarrow \mathbf{R}$ is linear
- ▶ then it can be expressed as $f(x) = a^T x$ for some a
- ▶ specifically: $a_i = f(e_i)$
- ▶ follows from

$$\begin{aligned} f(x) &= f(x_1 e_1 + x_2 e_2 + \cdots + x_n e_n) \\ &= x_1 f(e_1) + x_2 f(e_2) + \cdots + x_n f(e_n) \end{aligned}$$

Affine functions

- ▶ a function that is linear plus a constant is called *affine*
- ▶ general form is $f(x) = a^T x + b$, with a an n -vector and b a scalar
- ▶ a function $f : \mathbf{R}^n \rightarrow \mathbf{R}$ is affine if and only if

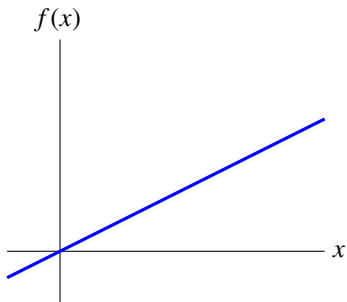
$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$$

holds for all α, β with $\alpha + \beta = 1$, and all n -vectors x, y

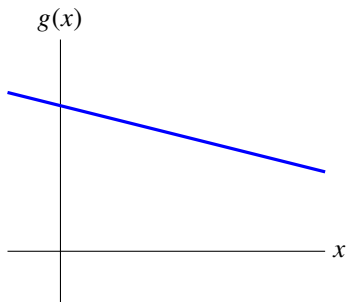
- ▶ sometimes (ignorant) people refer to affine functions as linear

Linear versus affine functions

f is linear



g is affine, not linear



Outline

Linear and affine functions

Taylor approximation

Regression model

First-order Taylor approximation

- ▶ suppose $f : \mathbf{R}^n \rightarrow \mathbf{R}$
- ▶ *first-order Taylor approximation* of f , near point z :

$$\hat{f}(x) = f(z) + \frac{\partial f}{\partial x_1}(z)(x_1 - z_1) + \cdots + \frac{\partial f}{\partial x_n}(z)(x_n - z_n)$$

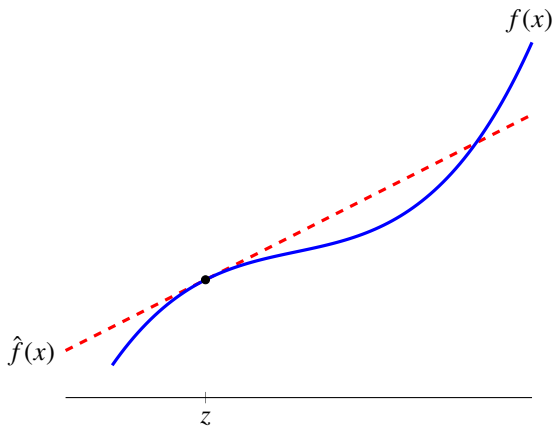
- ▶ $\hat{f}(x)$ is very close to $f(x)$ when x_i are all near z_i
- ▶ \hat{f} is an affine function of x
- ▶ can write using inner product as

$$\hat{f}(x) = f(z) + \nabla f(z)^T (x - z)$$

where n -vector $\nabla f(z)$ is the *gradient* of f at z ,

$$\nabla f(z) = \left(\frac{\partial f}{\partial x_1}(z), \dots, \frac{\partial f}{\partial x_n}(z) \right)$$

Example



Outline

Linear and affine functions

Taylor approximation

Regression model

Regression model

- ▶ *regression model* is (the affine function of x)

$$\hat{y} = x^T \beta + v$$

- ▶ x is a feature vector; its elements x_i are called *regressors*
- ▶ n -vector β is the *weight vector*
- ▶ scalar v is the *offset*
- ▶ scalar \hat{y} is the *prediction*
(of some actual outcome or *dependent variable*, denoted y)

Example

- ▶ y is selling price of house in \$1000 (in some location, over some period)
- ▶ regressor is

$$x = (\text{house area, \# bedrooms})$$

(house area in 1000 sq.ft.)

- ▶ regression model weight vector and offset are

$$\beta = (148.73, -18.85), \quad v = 54.40$$

- ▶ we'll see later how to guess β and v from sales data