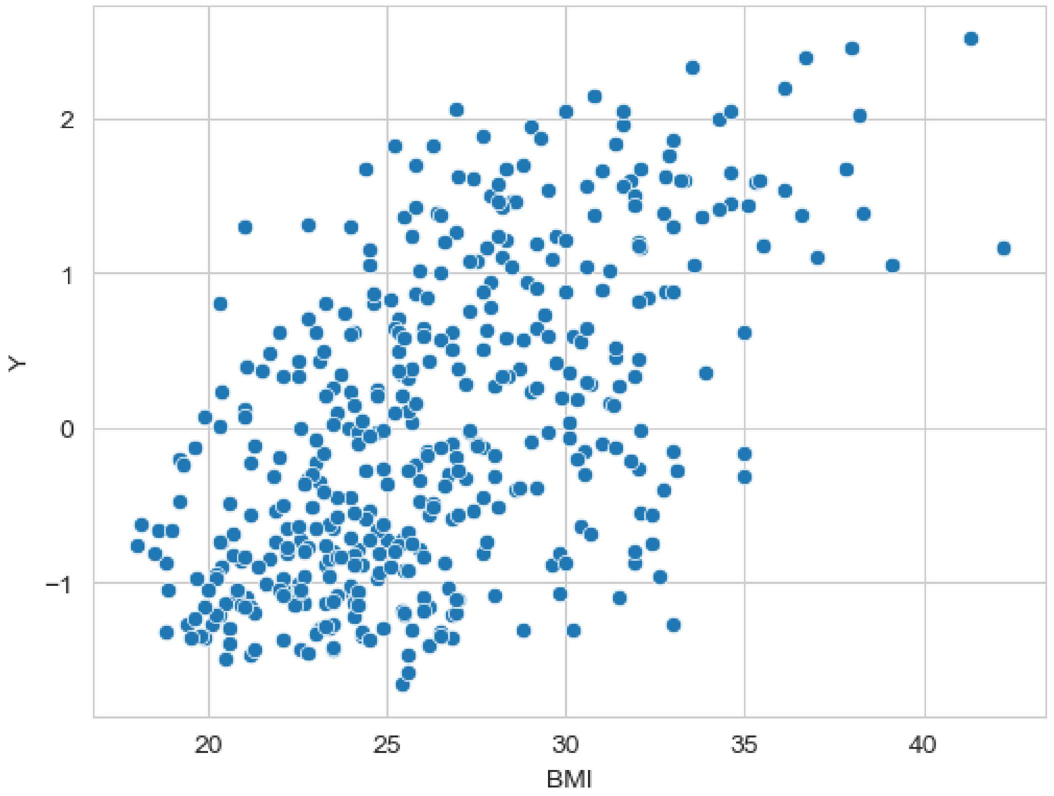
regression



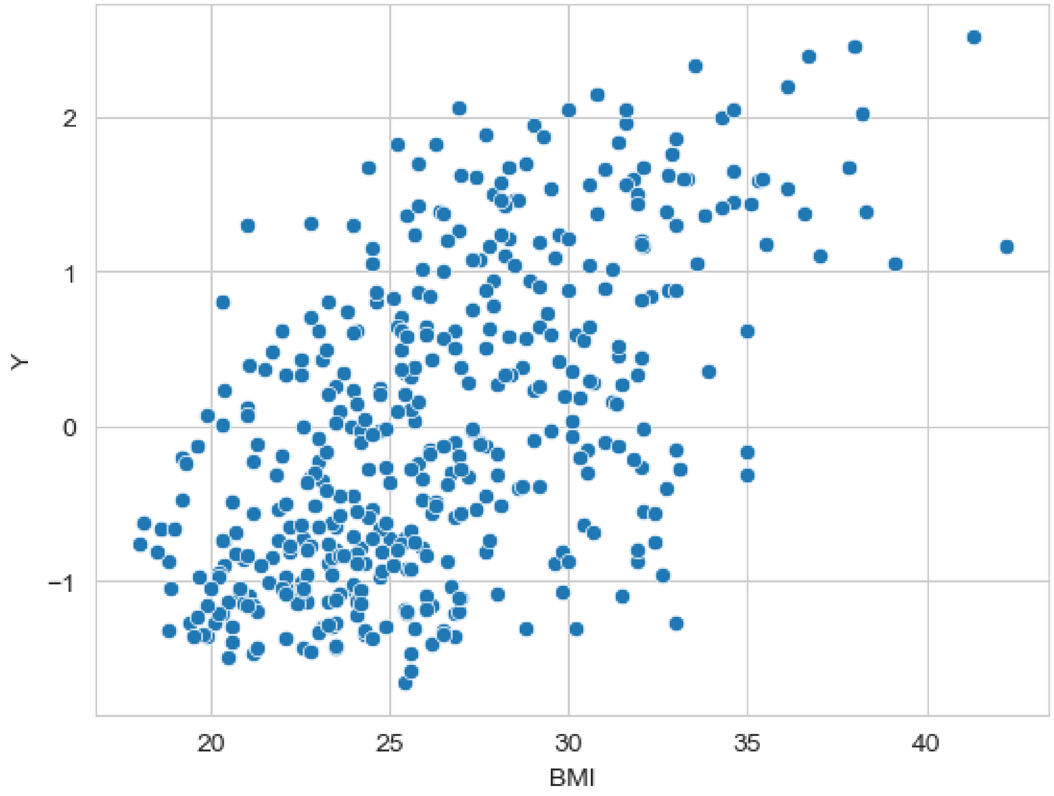
Can BMI explain Y?

Can BMI predict Y?

How does Y vary with BMI?

Regression is a general term for **modeling the relationship** between a **label** (a.k.a. dependent variable or target) and one or more **features** (a.k.a. the attributes, the independent or explanatory variable(s)).

linear regression



We need to make

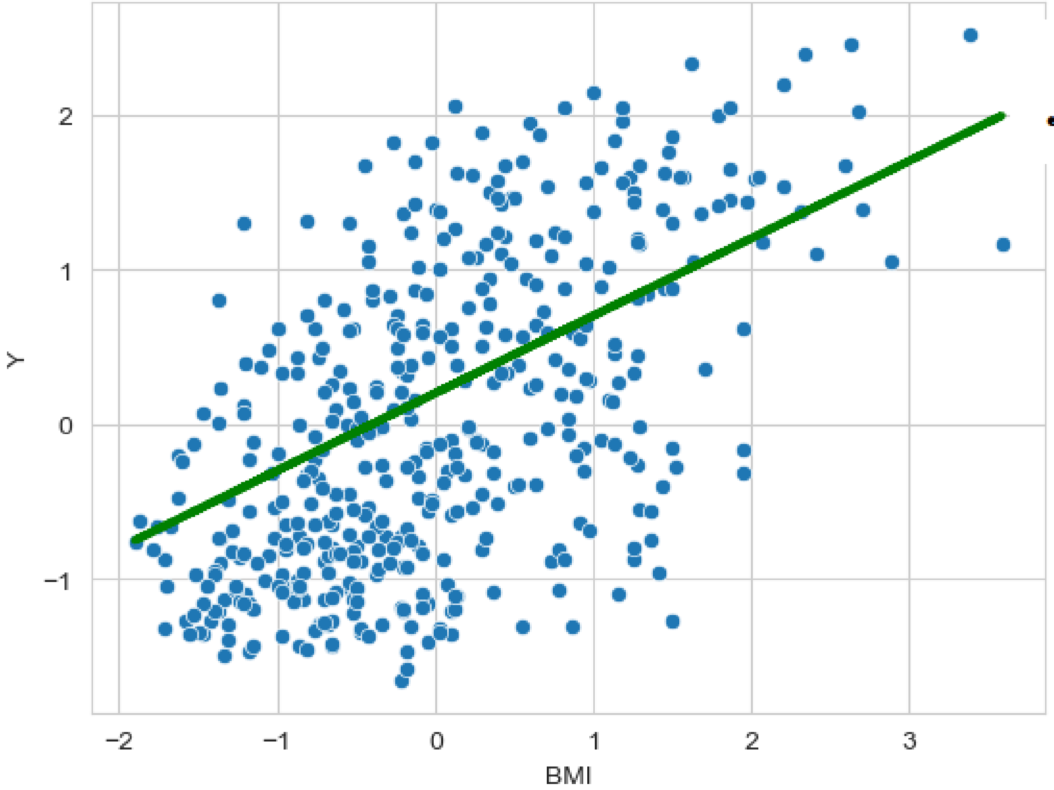
assumptions *linear relationship*

about the

model *linear model*

that generated the data.

linear regression



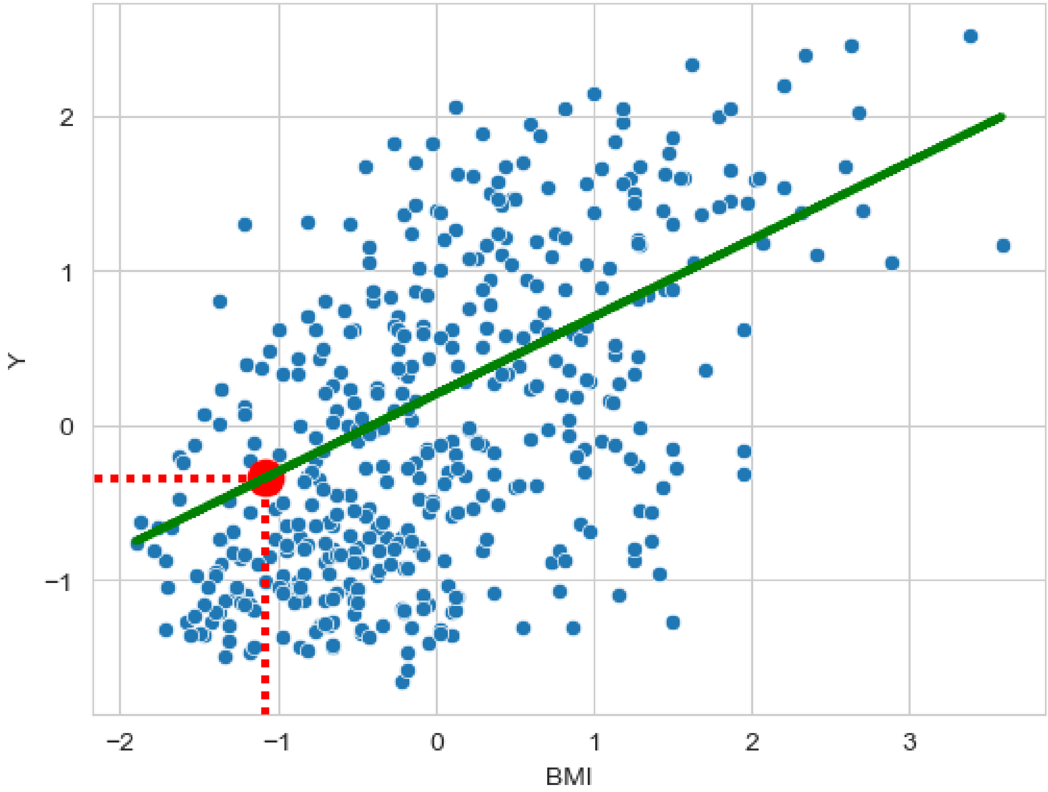
$$f(x) = 0.5 \cdot x + 0.2 = ax + b$$

a and *b* are **model parameters**

a is the **slope** to the line (direction)

b is the **intercept** or bias (position)

linear regression: prediction



BMI(70kg, 1.8 meter) = 21.6

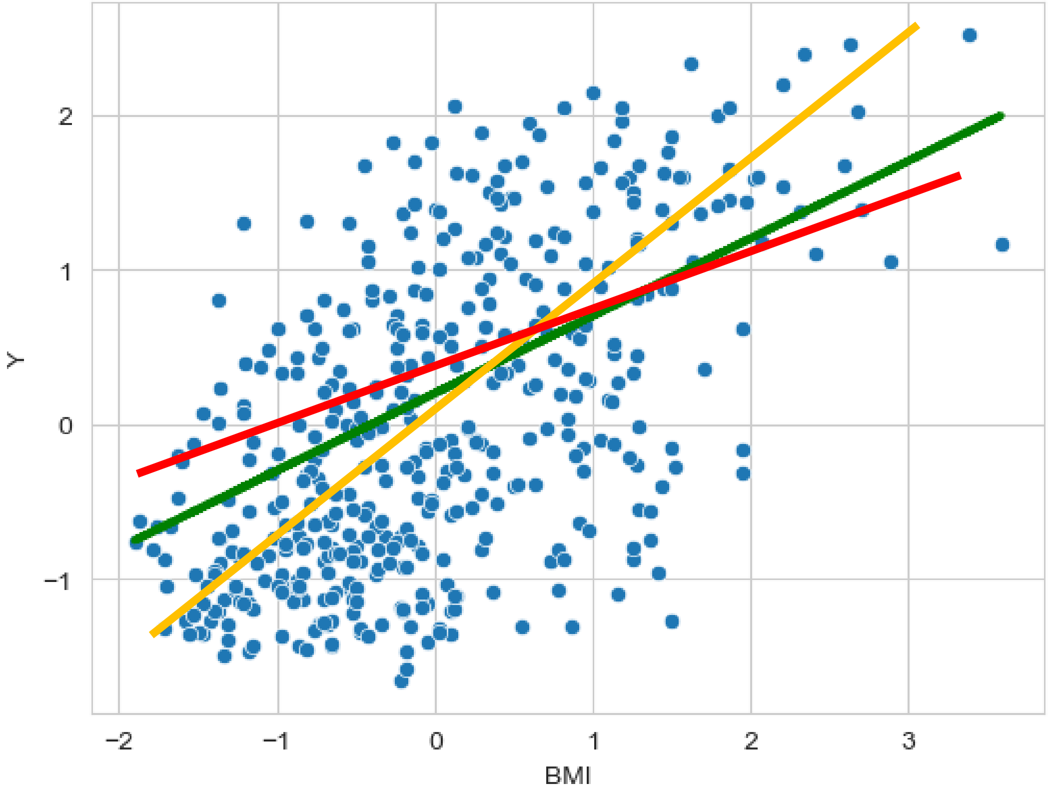
scaled data!

scaled BMI = -1.08

predicted value = -0.34

generalization!!

linear regression: fitting



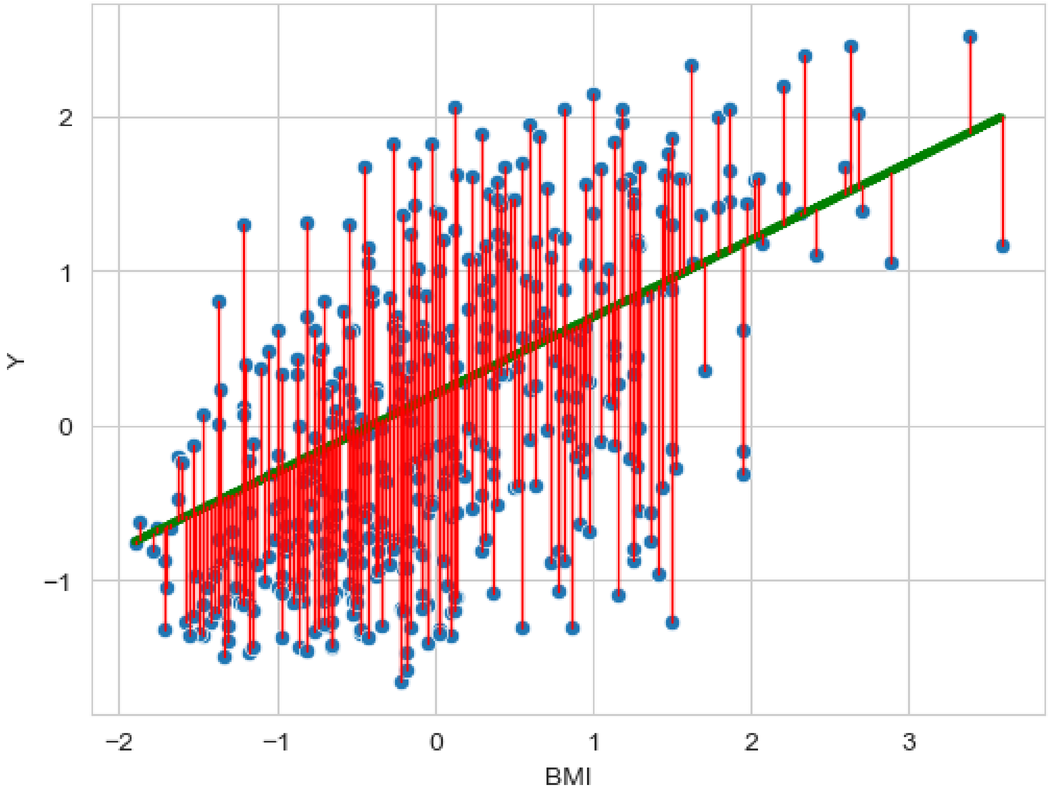
What values for *a* and *b* fit the data best?

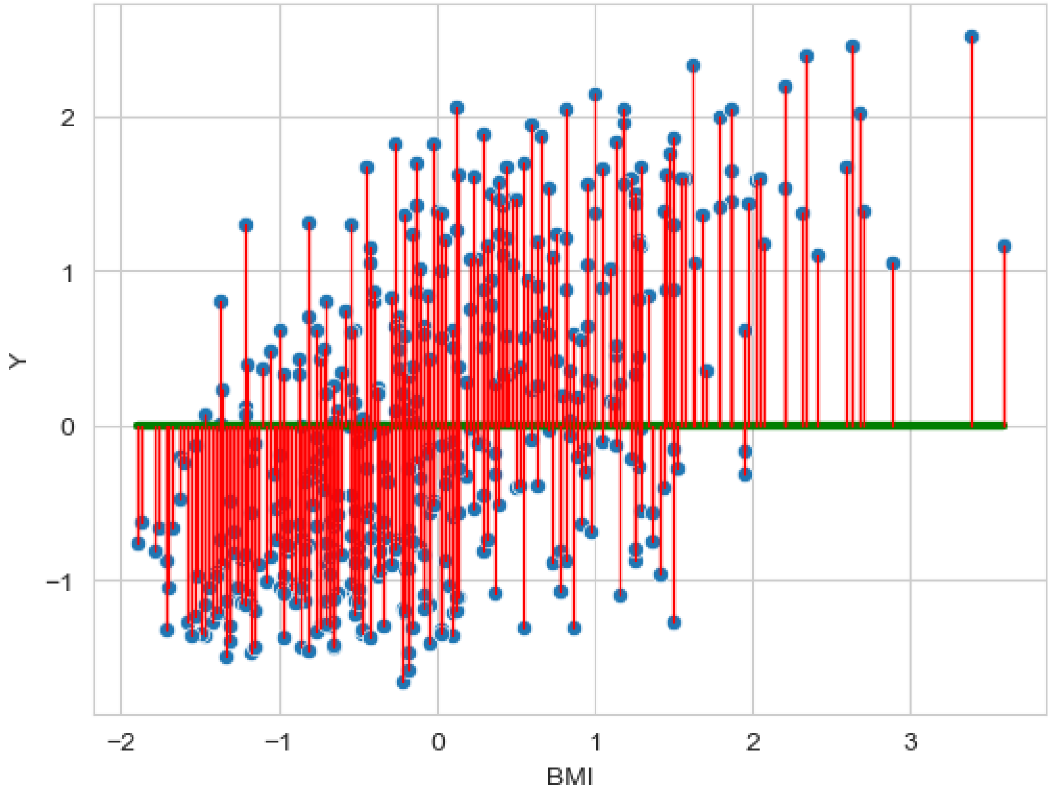
How can we evaluate them?

linear regression: fitting

$$SS_{error} = \sum_{i=1}^{n} (y^{(i)} - f(x^{(i)}))^2$$

$$SS_{total} = \sum_{i=1}^{n} (y^{(i)} - \bar{y})^2$$

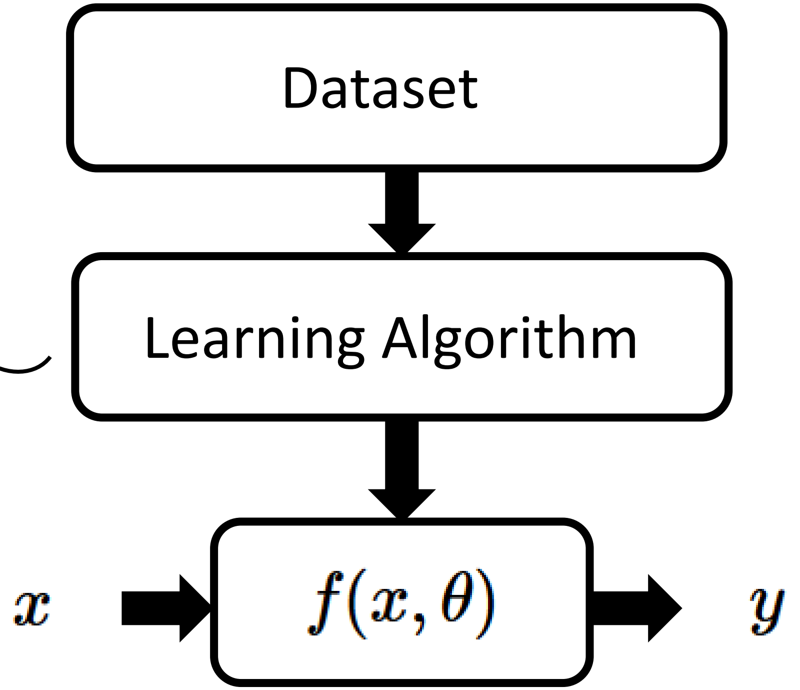




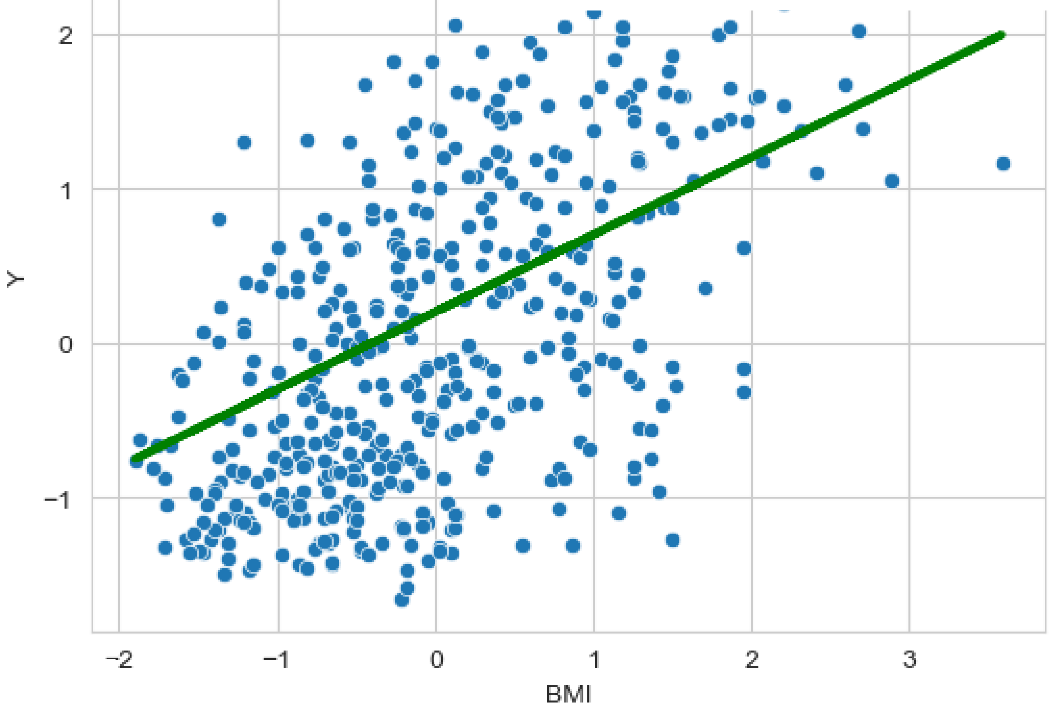
$$R^2 = 1 - \frac{SS_{error}}{SS_{total}} = 0.296$$

linear regression: formalized

linear regression



$$f(x, \theta) = \theta_0 \cdot \theta_1 x$$

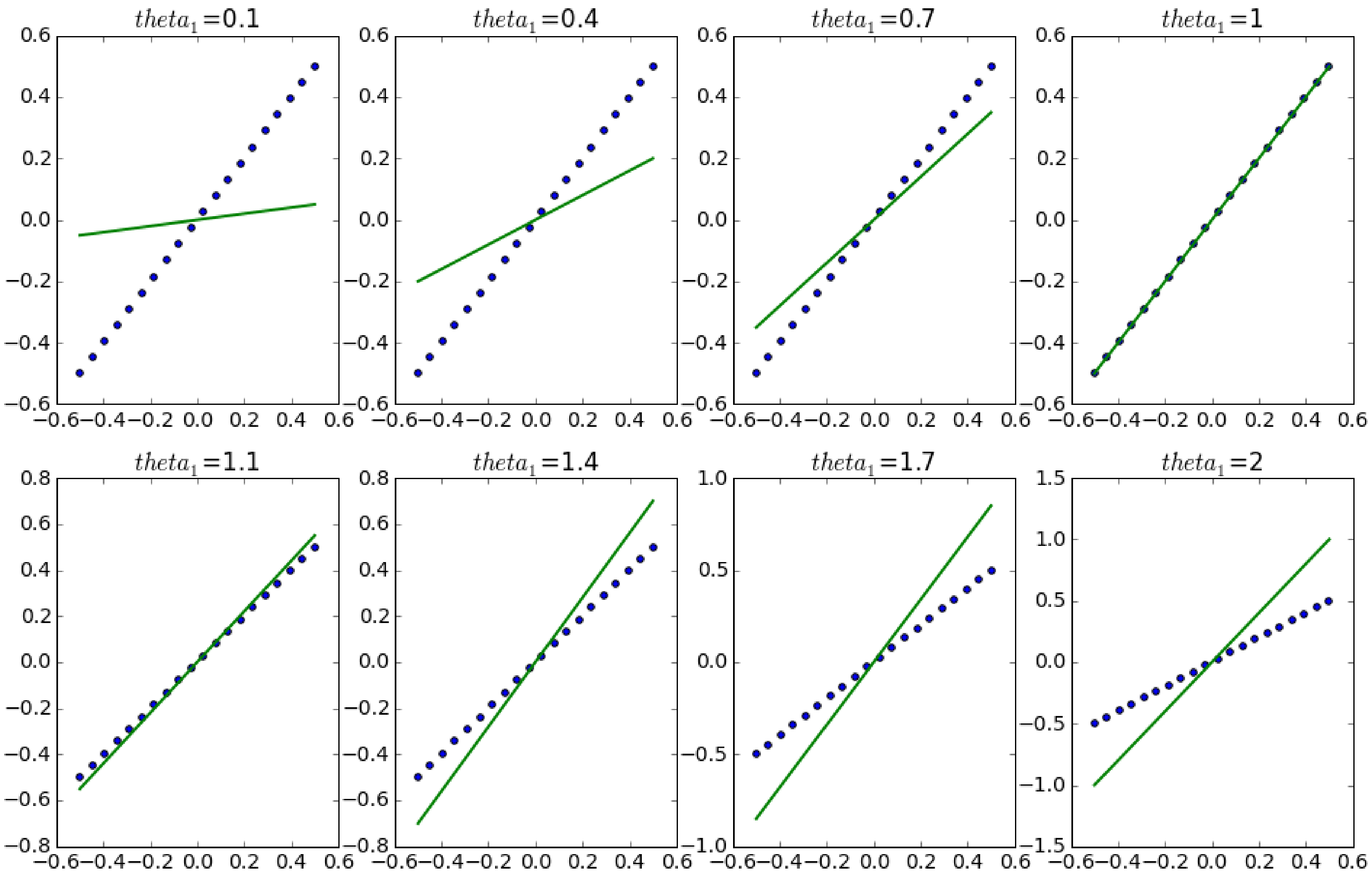


$$f(x, \theta) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_m x_m = \theta' x$$

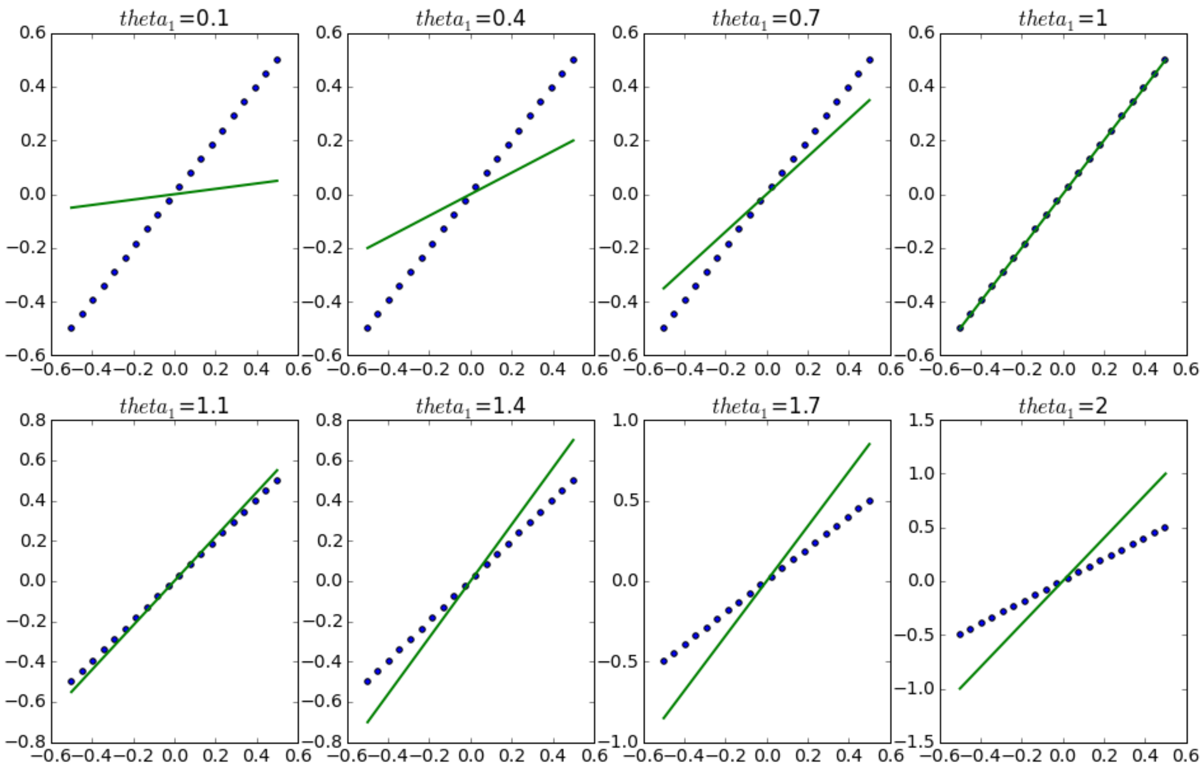
linear regression: cost (loss function)

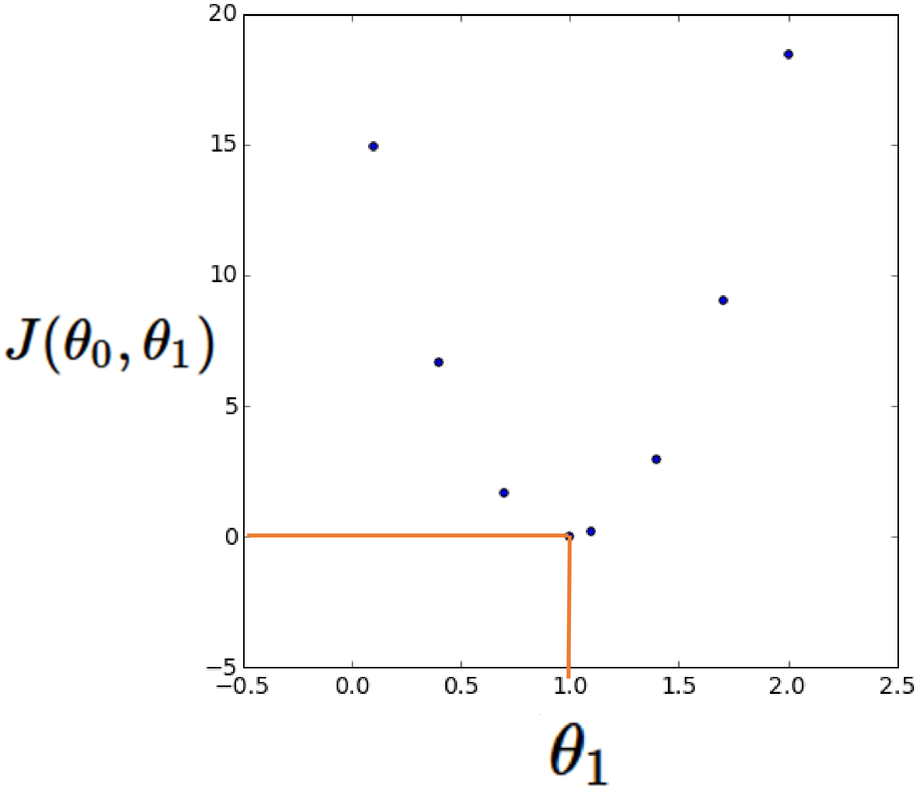
$$f(x) = x$$

$$(\theta_1 = 1, \theta_0 = 0)$$



linear regression: cost (loss function)





$$J(\theta_0, \theta_1) = \frac{1}{2n} \sum_{i=1}^{n} (f(x^{(i)}, \theta) - y^{(i)})^2$$

linear regression

Fit a linear model

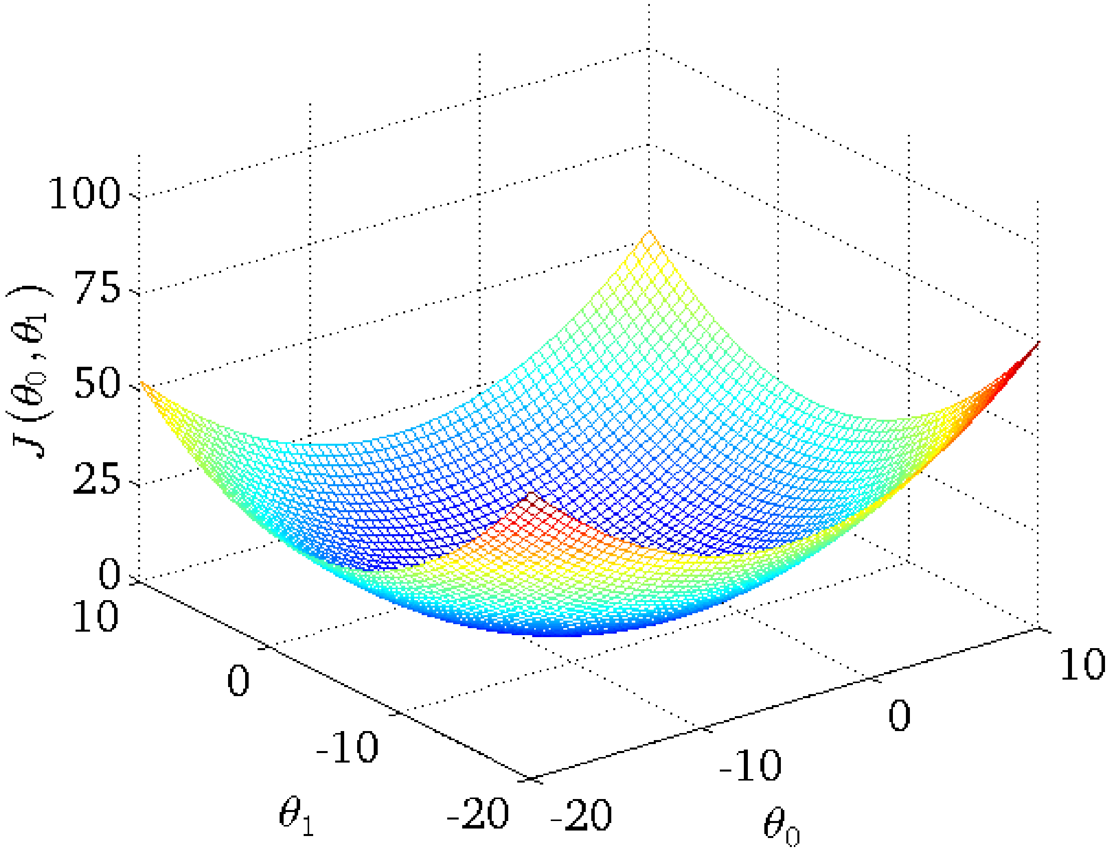
$$f(x,\theta) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_m x_m = \theta' x$$

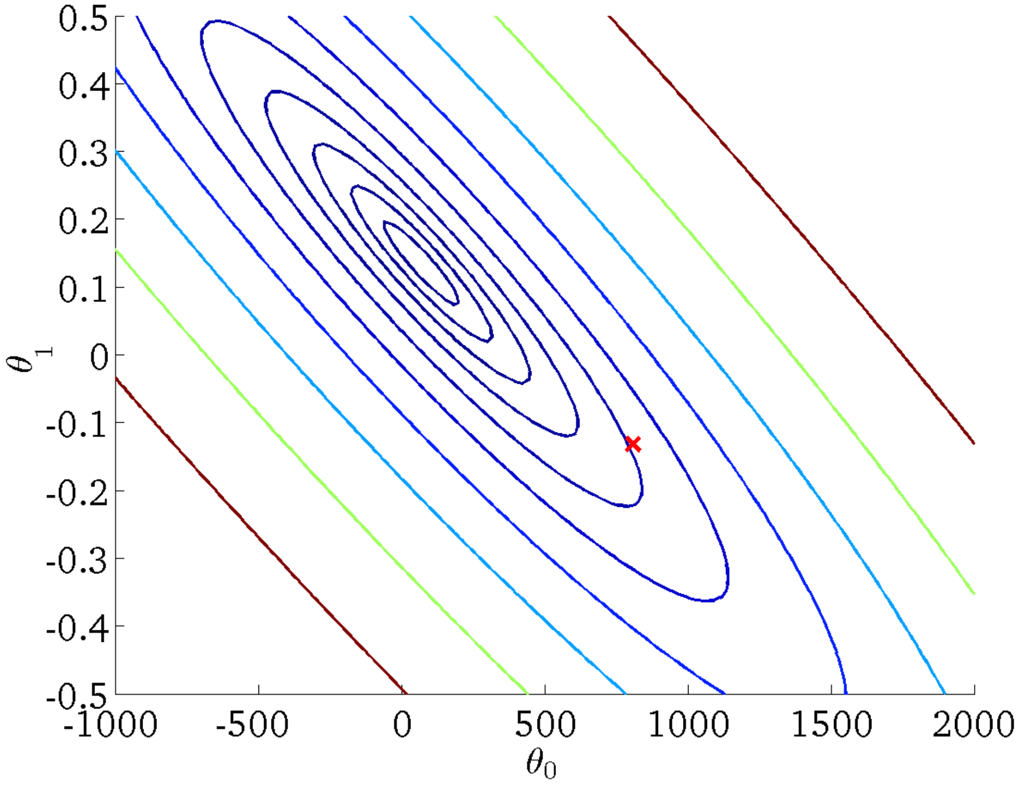
to the data set such that the cost function

$$J(\theta) = \frac{1}{2n} \sum_{i=1}^{n} (f(x^{(i)},\theta) - y^{(i)})^2$$

is minimal.

linear regression: cost (loss function)



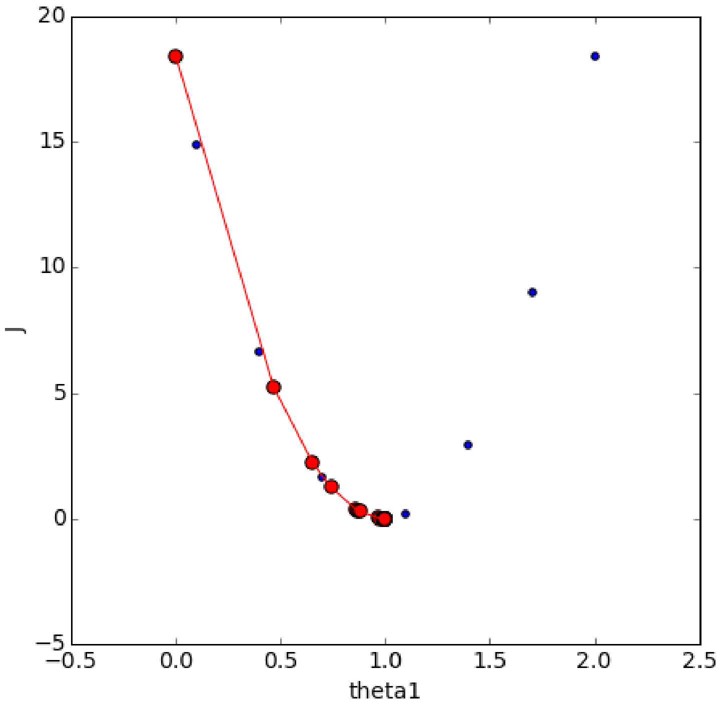


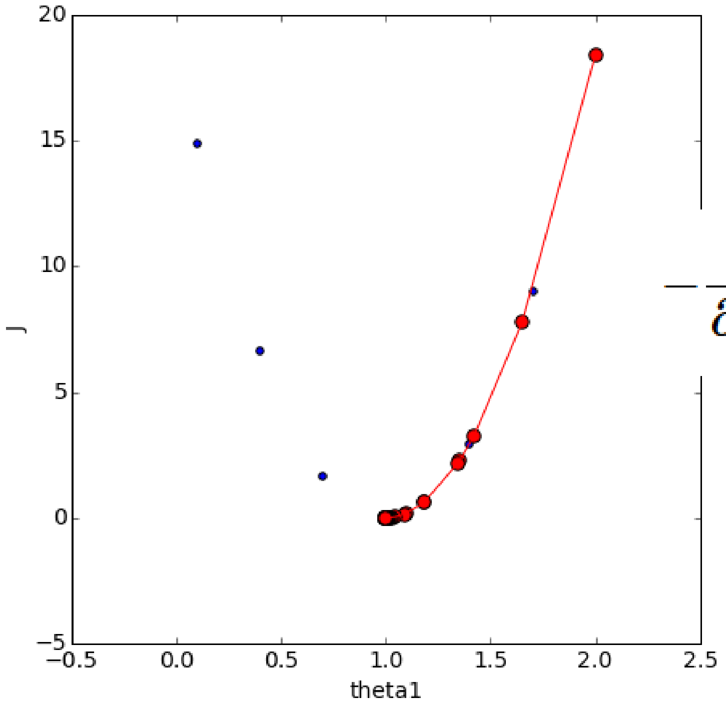
gradient descent

1. start with some random initialization of θ, i.e. a randomly chosen model

2. increment or decrement the value(s) of θ slightly such that $J(\theta)$ is reduced

3. repeat step 2. until we observe convergence





$$-\frac{\partial}{\partial\theta_j}J(\theta)$$

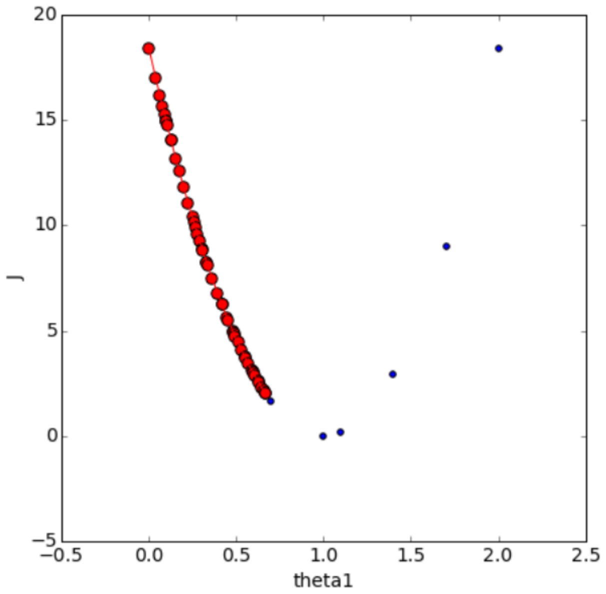
gradient descent: learning rate

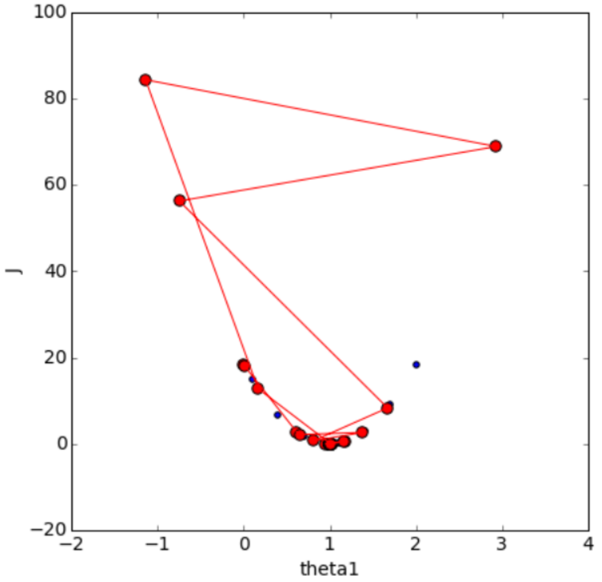
$$\theta_0 := \theta_0 - \alpha \frac{1}{n} \sum_{i=1}^{n} (f(x^{(i)}, \theta) - y^{(i)}) x_0^{(i)}$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{n} \sum_{i=1}^{n} (f(x^{(i)}, \theta) - y^{(i)}) x_1^{(i)}$$

$$\theta_2 := \theta_2 - \alpha \frac{1}{n} \sum_{i=1}^{n} (f(x^{(i)}, \theta) - y^{(i)}) x_2^{(i)}$$

$$\cdots$$





linear regression

model: $f(x, \theta) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_m x_m$

cost function: $J(\theta) = \frac{1}{2n} \sum_{i=1}^{n} (f(x^{(i)}, \theta) - y^{(i)})^2$

Goal: $\underset{J(\theta)}{\text{minimize}} \quad J(\theta)$

Learning: 1. start with some θ
2. change θ to reduce $J(\theta)$
3. repeat 2. until convergence

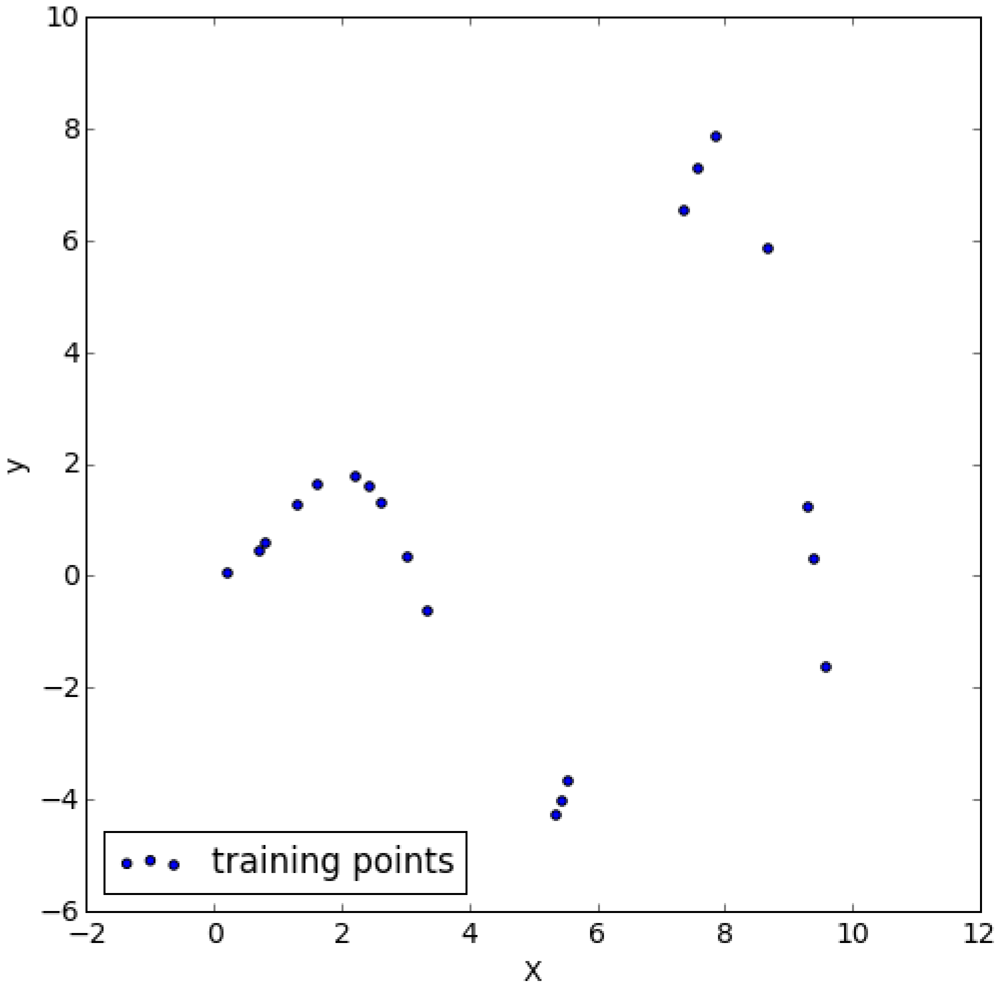
$$\theta_0 := \theta_0 - \alpha \frac{1}{n} \sum_{i=1}^{n} (f(x^{(i)}, \theta) - y^{(i)}) x_0^{(i)}$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{n} \sum_{i=1}^{n} (f(x^{(i)}, \theta) - y^{(i)}) x_1^{(i)}$$

$$\theta_2 := \theta_2 - \alpha \frac{1}{n} \sum_{i=1}^{n} (f(x^{(i)}, \theta) - y^{(i)}) x_2^{(i)}$$

...

non-linear regression



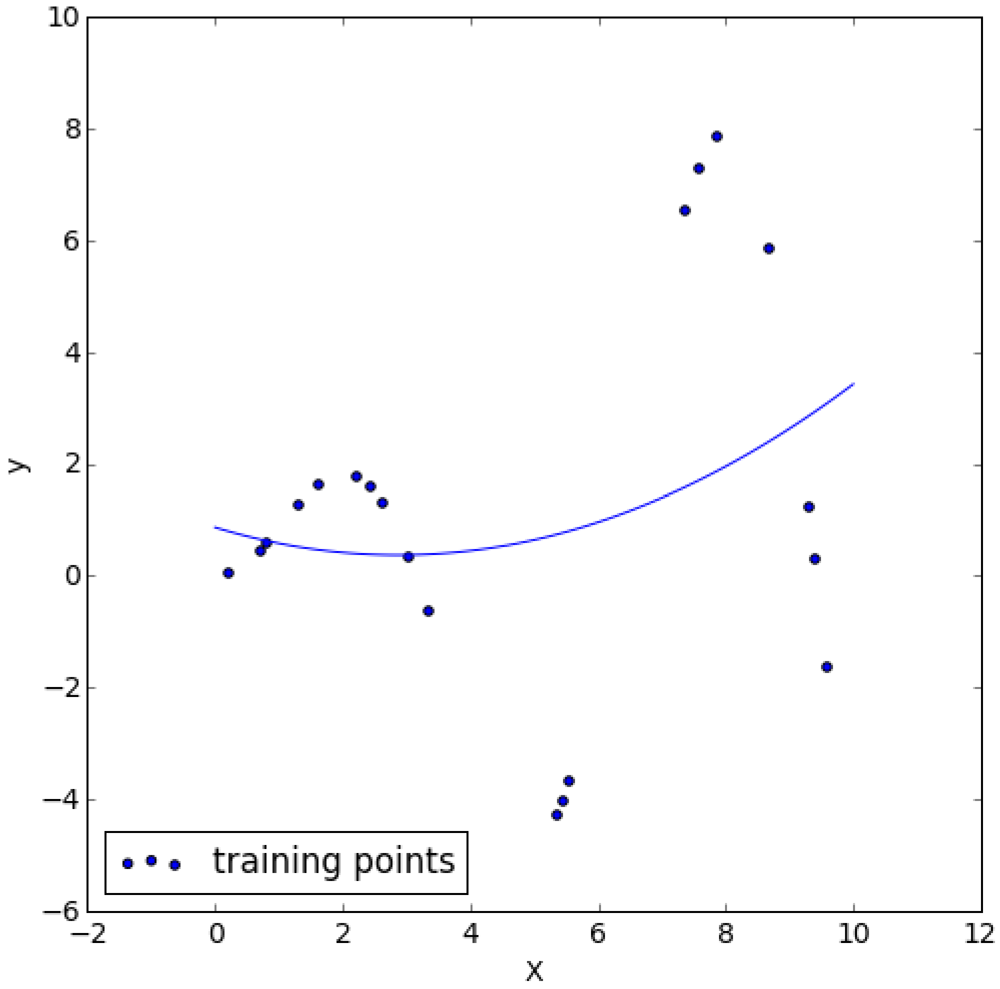
Does *y* vary linearly with *X*?

Can we fit a non-linear model? Yes

Or we could add polynomial transformations of the features.

$$f(x, \theta) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_1^2$$

non-linear regression

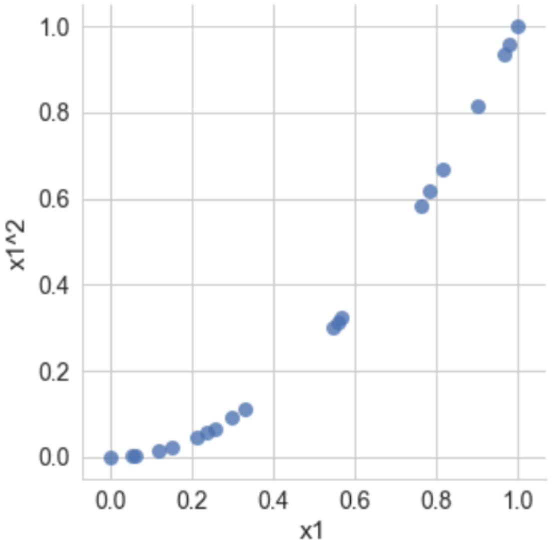


Does *y* vary linearly with *X*?

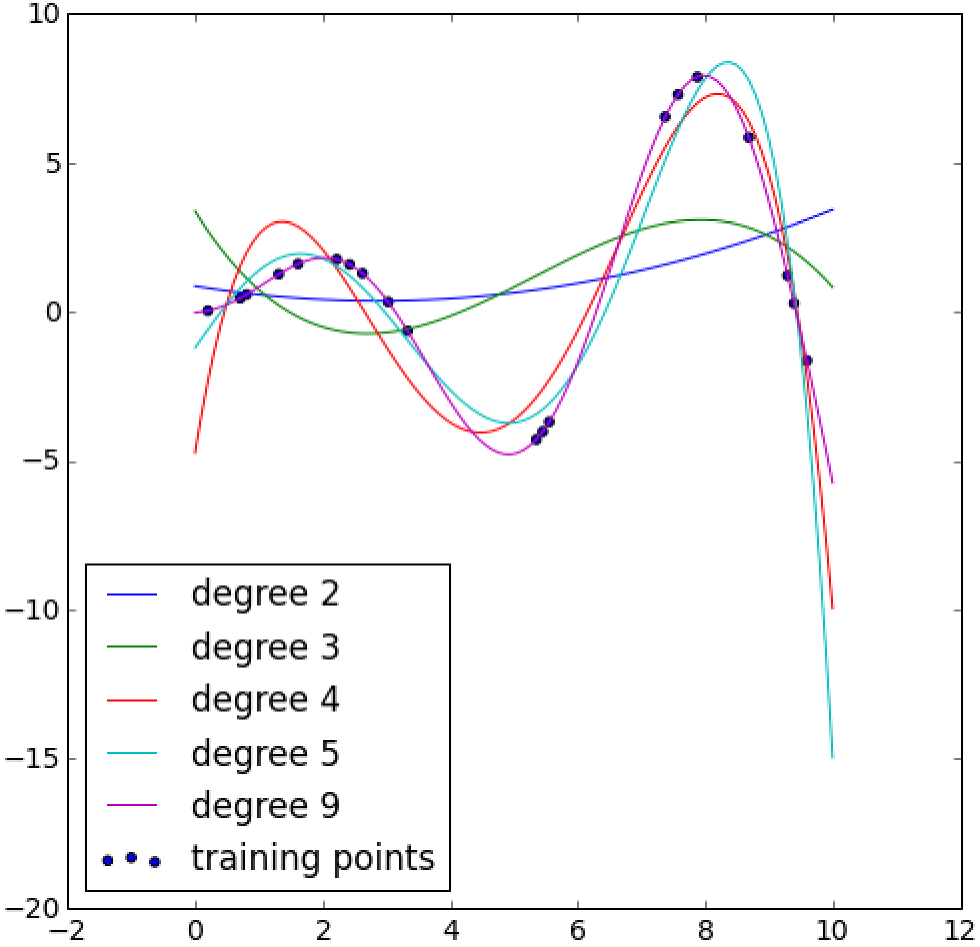
Can we fit a non-linear model? Yes

Or we could add polynomial transformations of the features.

$$f(x,\theta) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_1^2$$



non-linear regression



$$f(x,\theta) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^3 + \theta_4 x_1^4 + \theta_5 x_1^5 + \theta_6 x_1^6 + \theta_7 x_1^7 + \theta_8 x_1^8 + \theta_9 x_1^9$$