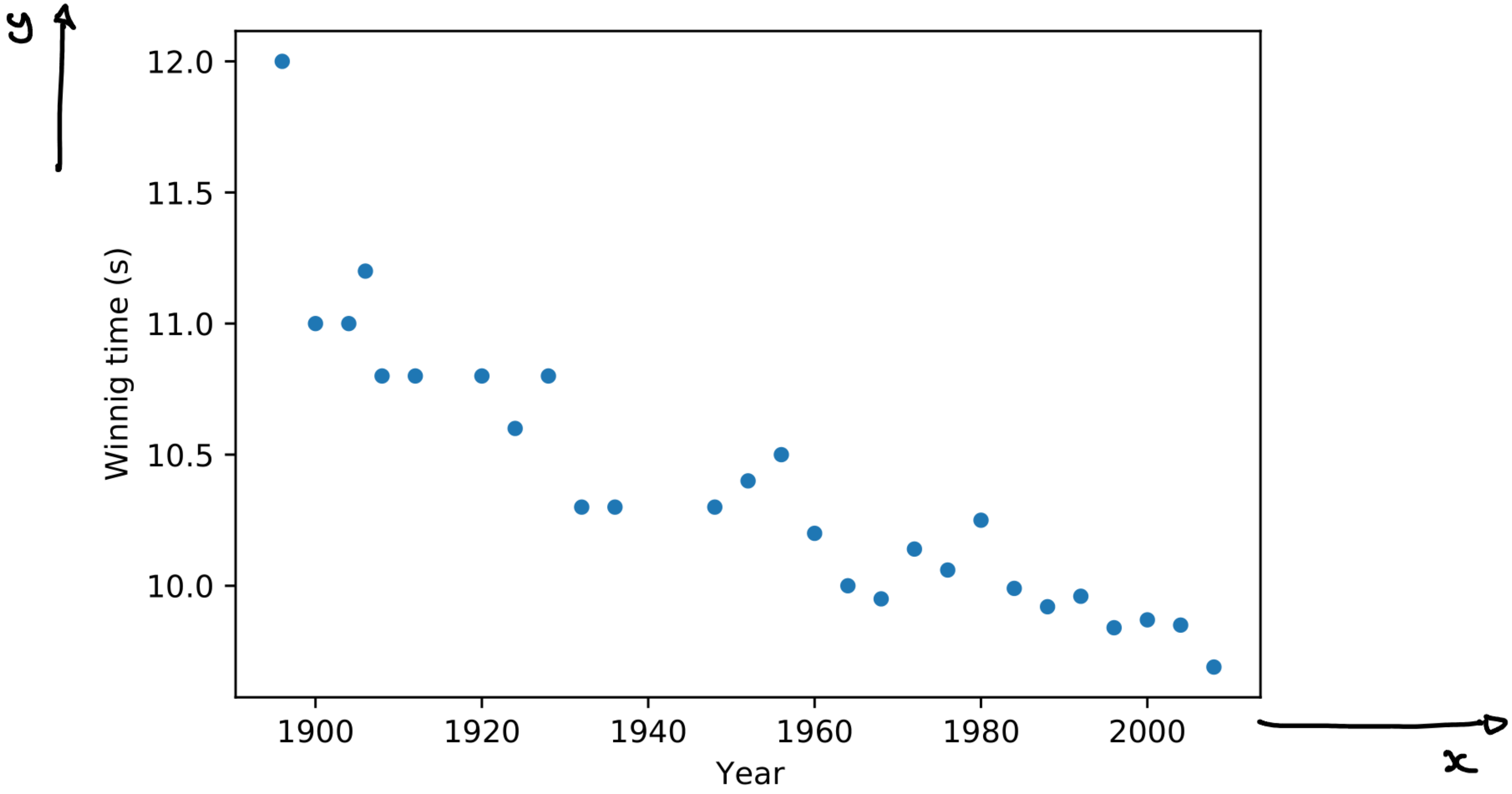
Simple linear regression

Herman Kamper

http://www.kamperh.com/

Winning 100-metre men's Olympic time from 1896 to 2008



Missing years: 1914, 1940, 1944

The model

A simple linear regression model predicts the output as a linear function of the input feature x:

$$f(x; w_0, w_1) = w_0 + w_1 x$$

We refer to w_0 and w_1 as the *parameters* of the model.
To choose w_0 and w_1, we are given a data set of previous input-output measurements:

$$\left\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots (x^{(N)}, y^{(N)})\right\}$$

I will sometimes just write this as:

$$\left\{(x^{(n)}, y^{(n)})\right\}_{n=1}^{N}$$

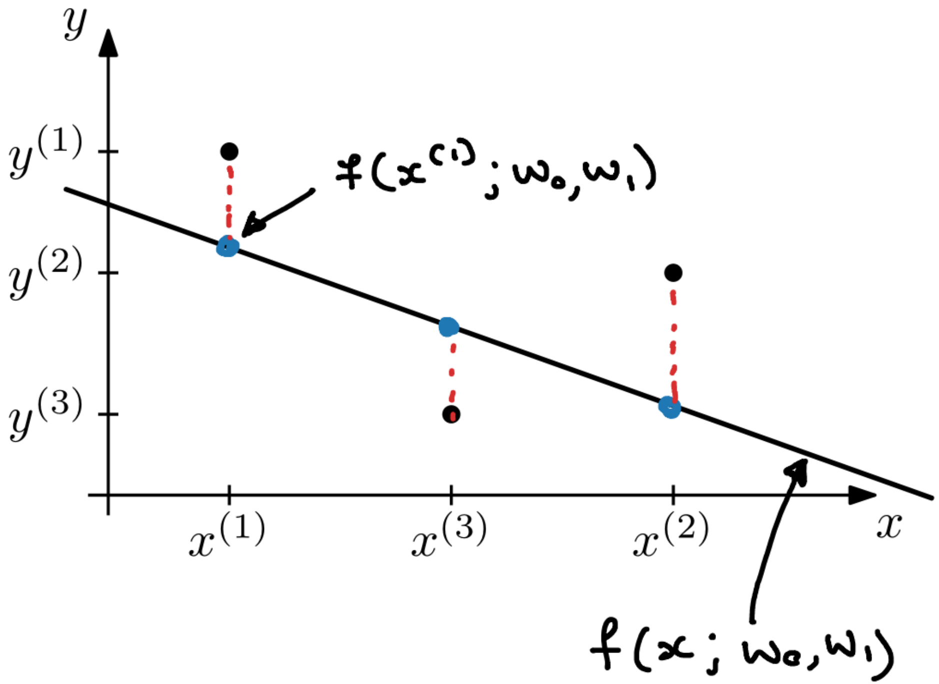
How do we choose w_0 and w_1 based on the data? We need some way to measure the "goodness" or "badness" of the parameters, given the data.

Loss function

(Sometimes called the cost function.)

How "good" is the fit of w_0, w_1 to this data?

$$J(w_0, w_1) = \sum_{n=1}^{N} \left(y^{(n)} - f(x^{(n)}; w_0, w_1) \right)^2$$



This is called the "squared loss" or the "residual sum of squares" (RSS).

Optimization

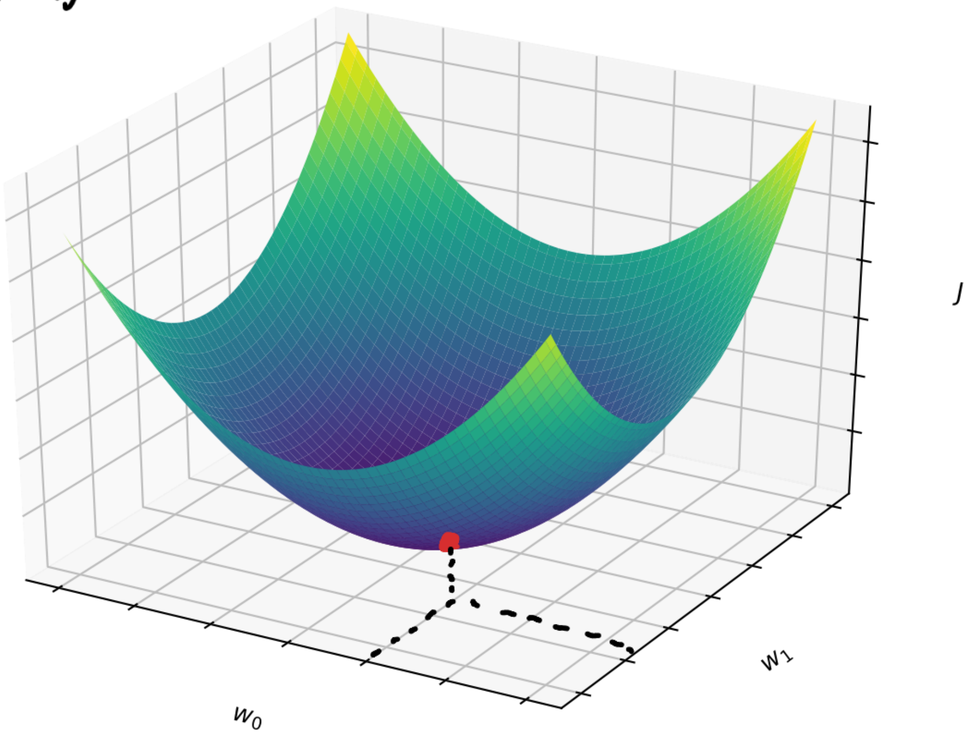
Want to find w_0, w_1 to minimise $J(w_0, w_1)$

$$\hat{w}_0, \hat{w}_1 = \underset{w_0, w_1}{\arg\min} \; J(w_0, w_1)$$

Strategy: Set $\frac{\partial J}{\partial w_0} = 0$ and $\frac{\partial J}{\partial w_1} = 0$

$$J(w_0, w_1) = \sum_{n=1}^{N} \left(y^{(n)} - f(x^{(n)}; w_0, w_1)\right)^2$$

$$= \sum_{n=1}^{N} \left(y^{(n)} - (w_0 + w_1 x^{(n)})\right)^2$$



$$J(w_0, w_1) = \sum_{n=1}^{N} \left(y^{(n)} - (w_0 + w_1 x^{(n)}) \right)^2$$

$$\frac{\partial J}{\partial w_0} = \sum_{n=1}^{N} \frac{\partial}{\partial w_0} \left(y^{(n)} - (w_0 + w_1 x^{(n)}) \right)^2$$

$$= \sum_{n=1}^{N} 2 \left(y^{(n)} - w_0 - w_1 x^{(n)} \right) \cdot (-1)$$

Set $\frac{\partial J}{\partial w_0} = 0$:

$$0 = \sum_{n=1}^{N} \cancel{2} \left(y^{(n)} - w_0 - w_1 x^{(n)} \right) \cancel{(-1)}$$

$$\sum_{n=1}^{N} w_0 = \sum_{n=1}^{N} y^{(n)} - w_1 \sum_{n=1}^{N} x^{(n)}$$

$$N w_0 = \sum_{n=1}^{N} y^{(n)} - w_1 \sum_{n=1}^{N} x^{(n)}$$

$$w_0 = \frac{1}{N} \sum_{n=1}^{N} y^{(n)} - w_1 \cdot \frac{1}{N} \sum_{n=1}^{N} x^{(n)}$$

$$\hat{w}_0 = \bar{y} - w_1 \cdot \bar{x} \qquad \ldots (1)$$

- Hat used to indicate particular value.
- Bar used to indicate estimated mean

$$\frac{\partial J}{\partial w_1} = \sum_{n=1}^{N} \frac{\partial}{\partial w_1} \left(y^{(n)} - w_0 - w_1 x^{(n)} \right)^2 \quad \text{[Use (1)]}$$

$$= \sum_{n=1}^{N} \frac{\partial}{\partial w_1} \left(y^{(n)} - \bar{y} + w_1 \bar{x} - w_1 x^{(n)} \right)^2$$

$$= \sum_{n=1}^{N} \frac{\partial}{\partial w_1} \left(y^{(n)} - \bar{y} - w_1 (x^{(n)} - \bar{x}) \right)^2$$

$$= \sum_{n=1}^{N} 2 \left(y^{(n)} - \bar{y} - w_1 (x^{(n)} - \bar{x}) \right) \cdot (-1) (x^{(n)} - \bar{x})$$

Set $\frac{\partial J}{\partial w_1} = 0$:

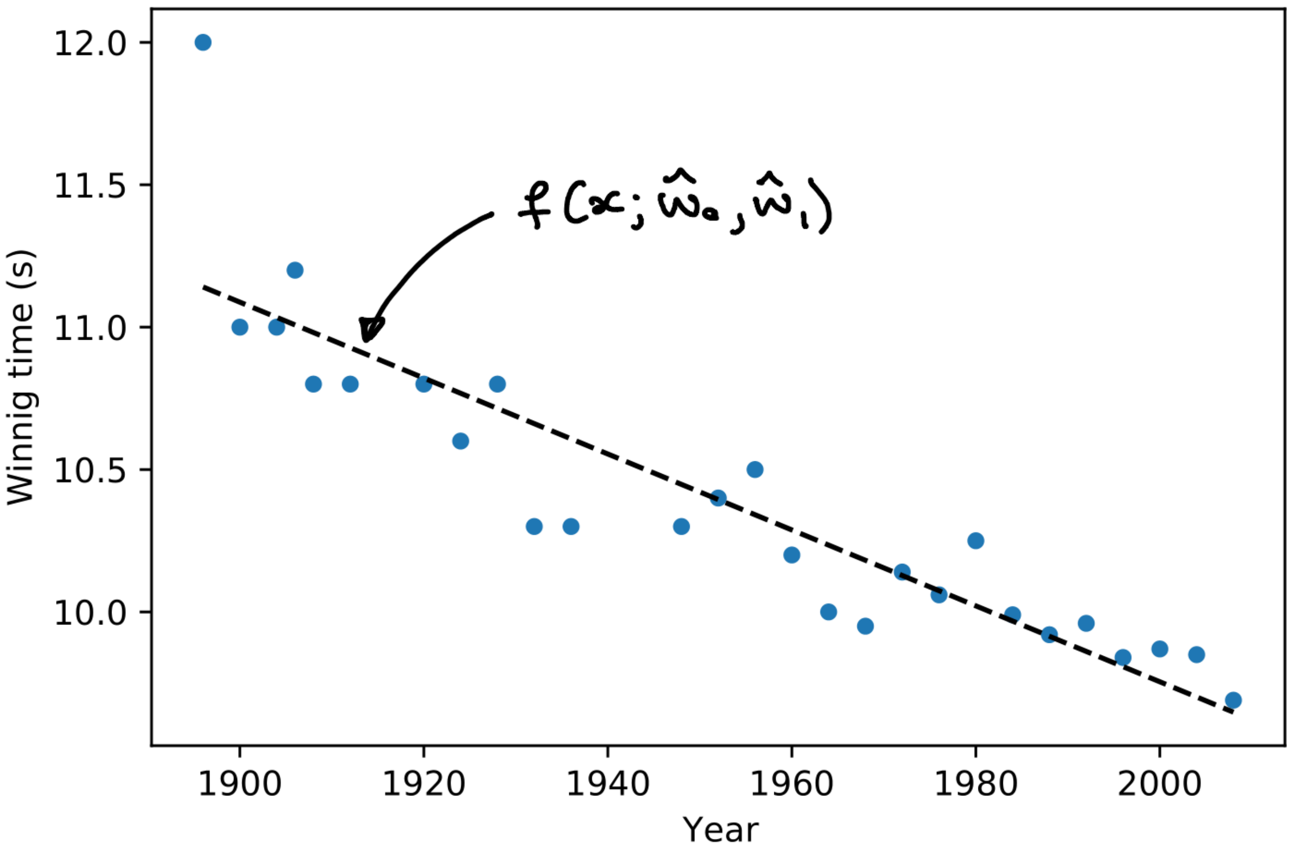
$$\hat{w}_1 = \frac{\sum_{n=1}^{N} (y^{(n)} - \bar{y})(x^{(n)} - \bar{x})}{\sum_{n=1}^{N} (x^{(n)} - \bar{x})^2}$$

These parameter settings are called the "least squares estimates."

We also need to show that there is one turning point in L corresponding to the minimum :

$$\frac{\partial^2 J}{\partial w_0^2} > 0 \quad \text{and} \quad \frac{\partial^2 J}{\partial w_1^2} > 0$$

Model fit



$f(x; \hat{w}_0, \hat{w}_1)$
12.0
11.5
11.0
10.5
10.0
Winnig time (s)
1900
1920
1940
1960
1980
2000
Year

Model predictions

- Estimated winning time in 1914: 10.901 s
- Estimated winning time in 2012: 9.595 s (actual time: 9.63 s)
- Estimated winning time in 2592: 1.863 s