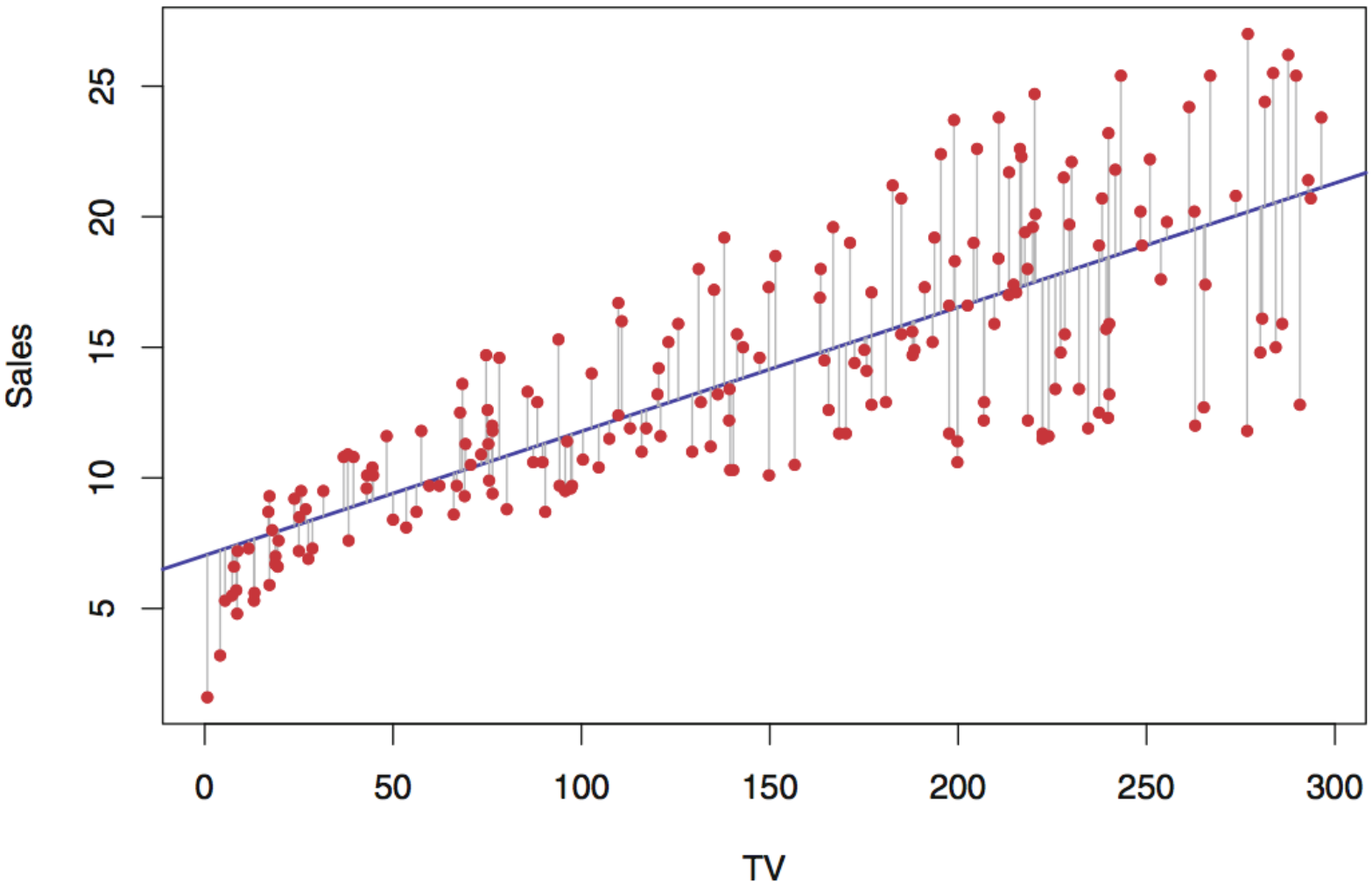
Least Squares Regression

Signal Data Science


Sales
5
10
15
20
25
0
50
100
150
200
250
300
TV

Why is linear regression important?

- Linear methods are simple and easy to calculate
- A surprising number of things are linear
- For many nonlinear things, linear methods get us pretty decent results
- Easily interpreted

Linearity in Nature

- "Complex nonlinear genetic systems certainly exist … However, quantitative differences *between* individuals … may be largely due to independent linear effects … As noted, **linear effects are the most readily evolvable in selection**, whereas nonlinear gadgets are more likely to be fragile to small changes. … [T]o first approximation, ***Biology = linear combinations of nonlinear gadgets***"
- From Steve Hsu's "On the genetic architecture of intelligence and other quantitative traits", **2014**

Least squares for regression

- Linear models can be fit with **many different cost functions**
- Method of least squares = minimize the sum of squared errors
- Why do we do this? Why not something else?

Gauss

- “Therefore, that will be the most probable system of values of the unknown quantities … in which the sum of the squares of the differences between the observed and computed values … is a minimum … **This principle, which promises to be of most frequent use in all applications of the mathematics to natural philosophy, must, everywhere, be considered an axiom** …”
- From Gauss’s “Theory of the motion of heavenly bodies”, **1809**

Gauss

- “It is very remarkable that the free movements, if they cannot coexist with the necessary conditions, are **modified by Nature in exactly the same way in which the calculating mathematician, according to the method of least squares, reconciles observations** which are connected to each other by necessary dependencies.”
- From Gauss’s “Werke”, Vol. 5, a paper on an extremal principle in mechanics

Assumptions

- Intuition: assume that **errors are normally distributed with mean 0 and identical variance & independent**
- Result: probability of observing our observations is maximized
- Maximum likelihood argument
- (Whiteboard math time!)
- Final theorem is even stronger, removing dependence on normality

Where do linear models fail?

- **Linear model:** 90%
- **4 layer CNN:** 99%

References

- W.K. Buhler, “Gauss: A Biographical Study”, Interchapter IX, 1979
- C.F. Gauss, “Theory of the Motion of the Heavenly Bodies”, Book II, Third Section, 1809
- C.F. Gauss, “Theory of the Combination of Observations Least Subject to Errors”, 1823
- M. Merriman, “On the History of the Method of Least Squares”, 1877