

CS 66: Machine Learning

Prof. Sara Mathieson

Spring 2019



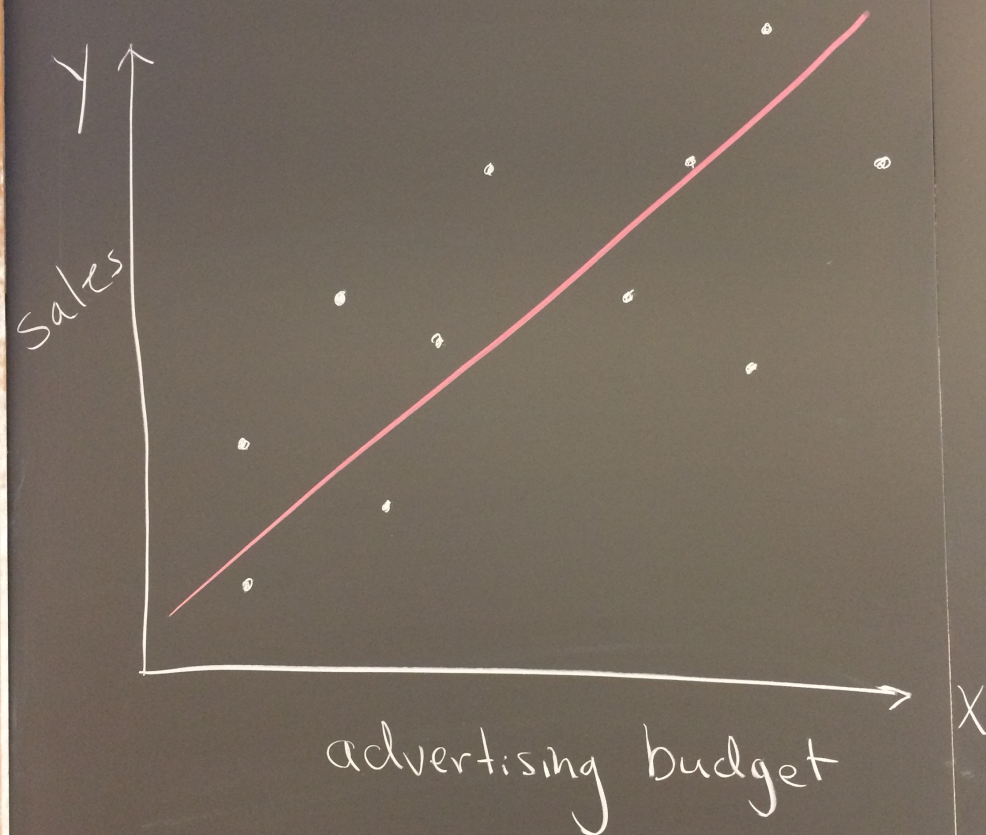
Outline for February 6

- Idea of *loss* for *regression* problems
- *Expected value*
- *Reducible* vs. *irreducible* error
- *Bias/variance* tradeoff
- Begin: *simple linear regression* (x has one feature)

Admin

- Lab 2 check-in **TODAY**

Example regression problem:
we might assume that as an advertising budget increases, sales will increase. Sales is a *continuous response variable*, so we are in a *regression* setting.



Regression Setup

Assume

$$y = f(x) + \varepsilon$$

error,
independent
of x , mean 0

$\hat{f}(x)$ is our estimate of f .

Why do we care?

prediction:

$$\hat{y} = \hat{f}(x)$$

mean 0 | loss function

$$l(y, \hat{y}) = (y - \hat{y})^2$$

compute: expected
value of loss

$$E[(y - \hat{y})^2] = ?$$

assume:

$$\hat{f} \neq x$$

are fixed.

weighted die (1, 2, 3, 4, 5, 6)

$$p_1 = \frac{1}{10}, p_2 = \frac{1}{10}, \dots, p_5 = \frac{1}{10}, p_6 = \frac{1}{2}$$

$$E[X] = \sum_{v \in \text{vals}(x)} p_v \cdot v$$

discrete
values

$$\begin{aligned} E[X] &= \frac{1}{10} \cdot 1 + \frac{1}{10} \cdot 2 + \dots + \frac{1}{2} \cdot 6 \\ &= \frac{1}{10} \cdot 15 + 3 \end{aligned}$$

$$E[X] = 4.5$$

weighted
average.

$$\rightarrow E[(y - \hat{y})^2] = E[(f(x) + \varepsilon - \hat{f}(x))^2]$$

$$= (f(x) - \hat{f}(x))^2 + E[(\varepsilon)^2] + 2E[\underbrace{\varepsilon(f(x) - \hat{f}(x))}_{\text{constant}}]$$

$$E[(y - \hat{y})^2] = \underbrace{(f(x) - \hat{f}(x))^2}_{\text{reducible error}} + \underbrace{\text{Var}(\varepsilon)}_{\text{irreducible error}}$$

$$E[\varepsilon] = 0$$

$$\text{Var}(X) = E[(X - E[X])^2]$$

$$E[7] = 7$$

)²]

$$E[\varepsilon^2] + 2E[\underbrace{\varepsilon(f(x) - \hat{f}(x))}_{\text{constant}}]$$

$$E[\varepsilon] = 0$$

$$\text{Var}(X) = E[(X - \mu)^2]$$

↑
mean
or expected
value

variable
error

Goals of Inference

- ① which features/explanatory variables/predictors $\{x\}$ are associated w/ response variable y ?
- ② what is the relationship between x & y ?
- ③ is linear model enough?

Accuracy

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2$$

mean squared

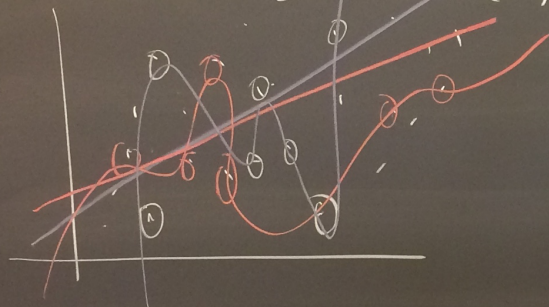
error

now
assume
not constant

* expected test MSE

$$E[(y - \hat{f}(x))^2] = \text{Var}(\hat{f}(x)) + [\text{Bias}(\hat{f}(x))]^2 + \text{Var}(\varepsilon)$$

• variance: amount \hat{f} would change if we used a different training set (X)



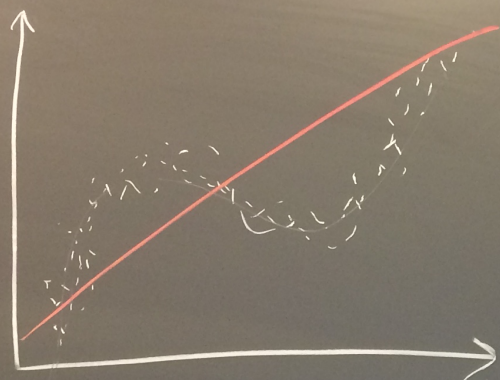
low flexibility
 \Rightarrow low variance

• bias: error introduced by approximating a real-life problem.

$\text{var}(\varepsilon)$

if

training



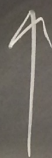
- low variance

- high bias

want: low variance
& low bias

but: as flexibility
increases:

• variance



• bias

