CS 383: Machine Learning

Prof Adam Poliak

Fall 2024

09/18/2024

Lecture 06

Announcements

HW02 is due Sunday night

- **Reading quiz: Thursday**
  - Duame 7.6 (2+ pages)
  - ISL 59-63 (4+ pages)

- Midterm 1: Thursday October 3rd

Outline

Expectation & Variance

Bias-Variance Tradeoff

Linear regression

Loss Function

$\ell\left(y, \hat{f}(x)\right)$ quantifies how far our prediction is from the true value

We want to minimize Expected Loss:

$$\mathbb{E}_{(x,y)}\left[\ell\left(y, \hat{f}(x)\right)\right]$$

Expected Values

Weight

$$\mathbb{E}[X] = \sum_{v \in X} p(X = v)\, v$$

Handout

$$Var(X) = \sigma_x^2 = \mathbb{E}[(X - \mu_x)^2]$$

$$SD(X) = \sigma_x = \sqrt{\mathbb{E}[(X - \mu_x)^2]}$$

Expected Value Rules

Additivity of Expectation

$$\mathbb{E}[X + Y] =$$

$$= \mathbb{E}[X] + \mathbb{E}[Y]$$

Linearity of Expectation

$$\mathbb{E}[\alpha X] =$$

$$= \alpha \mathbb{E}[X]$$

https://prob140.org/textbook/content/Chapter_08/04_Additivity.html

Expected Values

Quick Check

Let X and Y be random variables on the same space, with $E(X) = 5$ and $E(Y) = 3$.

(a) Find $E(X - Y)$.

(b) Find $E(2X - 8Y + 7)$.

https://prob140.org/textbook/content/Chapter_08/04_Additivity.html

Expected loss

Read it as (x,y) has distribution $\mathcal{D}$

$$(x, y) \sim \mathcal{D}$$

Probability of (x,y) occurring

loss

Expected Loss

$$\sum_{x,y \in D} D(x, y)\ \ell\left(y, \hat{f}(x)\right)$$

$$\frac{1}{n}\sum_{i}^{n} \ell\left(y_i, \hat{f}(x_i)\right)$$

Mean Squared Error

Squared error:
$\ell(y, \hat{y}) = (y - \hat{y})^2$

Mean Square Error (MSE):
$$\frac{1}{n}\sum_{i}^{n}(y_i - \widehat{y_i})^2$$

$$\mathbb{E}[MSE] = \frac{1}{n}\sum_{i}^{n}\mathbb{E}[(y_i - \widehat{y_i})^2]$$

Mean Squared Error

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

$$= \mathbb{E}\left[\left(y-\hat{f}\right)^2\right]$$

error ε

model issue

$$= \mathbb{E}\left[\left(y-f+f-\hat{f}\right)^2\right]$$

$$= Var(\varepsilon) + \mathbb{E}\left[\left(f-\hat{f}\right)^2\right]$$

Mean Squared Error

$$\mathbb{E}[(y-\hat{y})^2] = Var(\varepsilon) + \mathbb{E}\left[\left(f - \hat{f}\right)^2\right]$$

$$\mathbb{E}\left[\left(f - \hat{f}\right)^2\right] = \mathbb{E}\left[\left(f - \mathbb{E}[\hat{f}] + \mathbb{E}[\hat{f}] - \hat{f}\right)^2\right]$$

$$= \mathbb{E}\left[\left(f - \mathbb{E}[\hat{f}]\right)^2 + 2\left(\hat{f} - \mathbb{E}[\hat{f}]\right)\left(\mathbb{E}[\hat{f}] - \hat{f}\right) + \left(\mathbb{E}[\hat{f}] - \hat{f}\right)^2\right]$$

$$= \mathbb{E}\left[\left(f - \mathbb{E}[\hat{f}]\right)^2\right] + \mathbb{E}\left[2\left(\hat{f} - \mathbb{E}[\hat{f}]\right)\left(\mathbb{E}[\hat{f}] - \hat{f}\right)\right] +. \mathbb{E}\left[\left(\mathbb{E}[\hat{f}] - \hat{f}\right)^2\right]$$

$$= var\left(\hat{f}(x)\right) \qquad + \quad 0 \qquad\qquad + \quad bias\left(\hat{f}\right)^2$$

$$\mathbb{E}[MSE] = var(\varepsilon) + var\left(\hat{f}\right) + bias\left(\hat{f}\right)^2$$

Mean Squared Error

$$\mathbb{E}[MSE] = bias(\hat{f})^2 + var(\hat{f}) + var(\varepsilon)$$

Bias: approximation error

error introduced by approximated a real-life problem

$$\hat{f} - f$$

Variance: estimation error

amount $\hat{f}$ would change if we trained on different data

Bias Variance Tradeoff

Assessing Model Accuracy

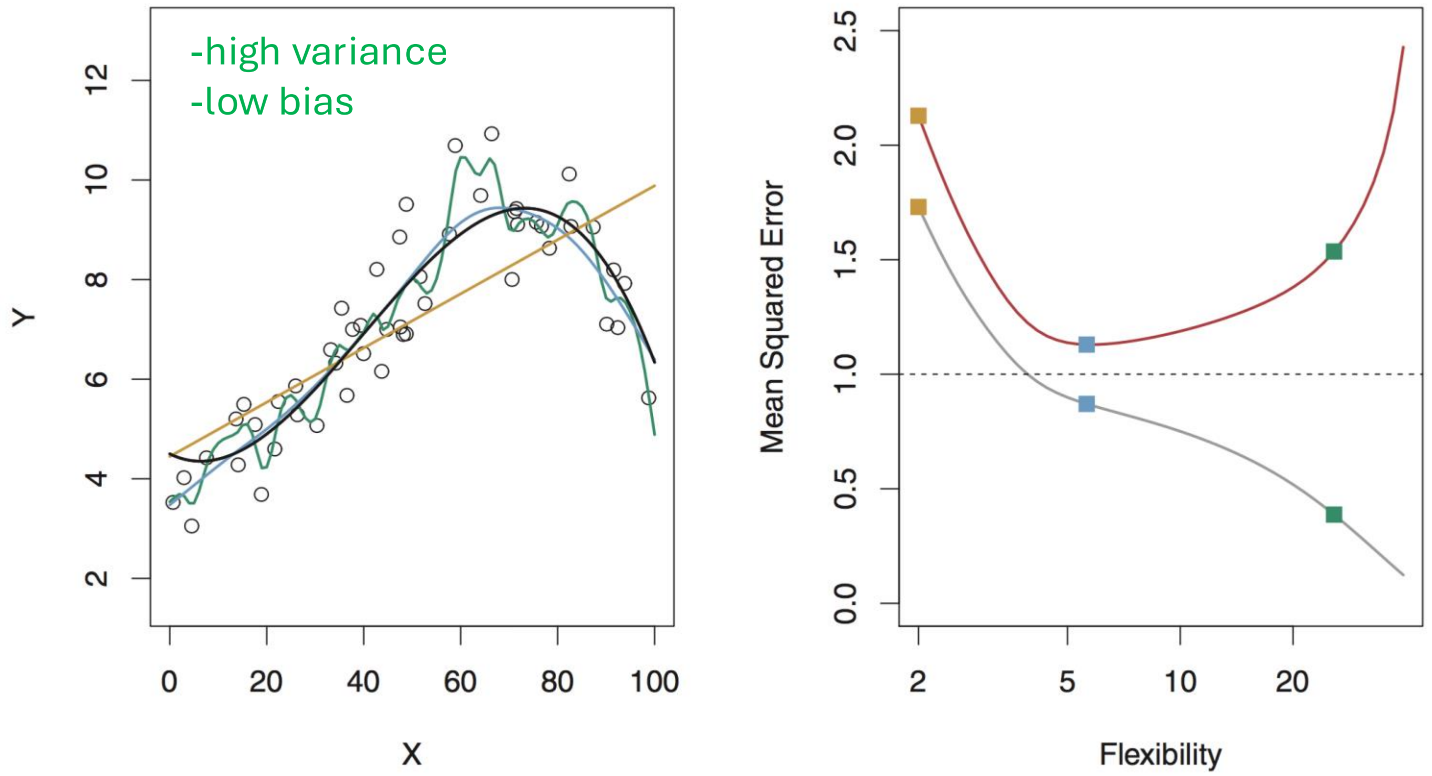


FIGURE 2.9. Left: *Data simulated from f, shown in black. Three estimates of f are shown: the linear regression line (orange curve), and two smoothing spline fits (blue and green curves).* Right: *Training MSE (grey curve), test MSE (red curve), and minimum possible test MSE over all methods (dashed line). Squares represent the training and test MSEs for the three fits shown in the left-hand panel.*

ISL, Fig 2.9

Outline

Continuous Features in Decision Trees

Learning problem so far + terminology

Bias-Variance Tradeoff

Linear regression

Goals of Inference

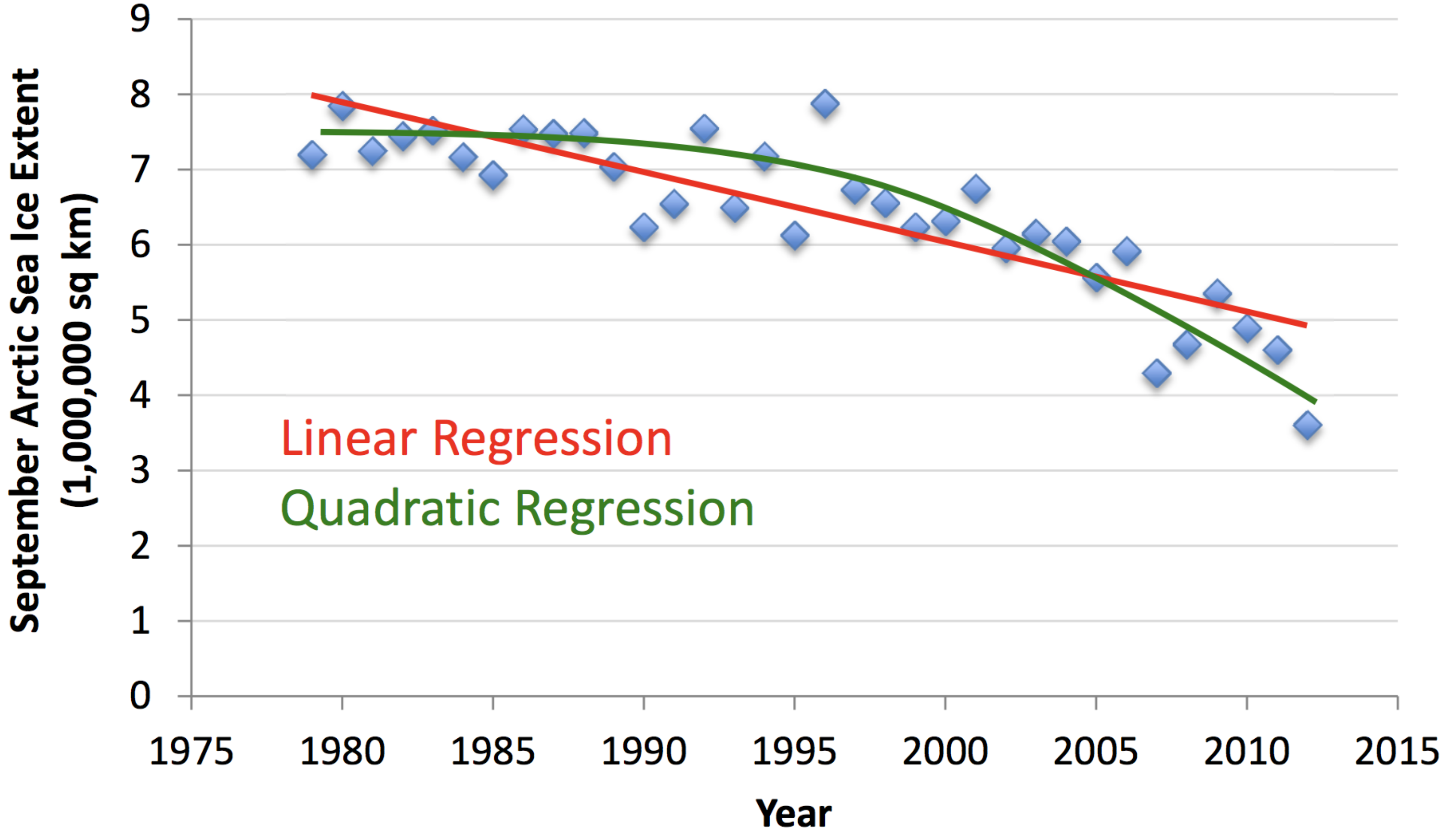
1) Which of the features/explanatory variables/predictors (x) are associated with the response variable (y)?

2) What is the relationship between x and y?

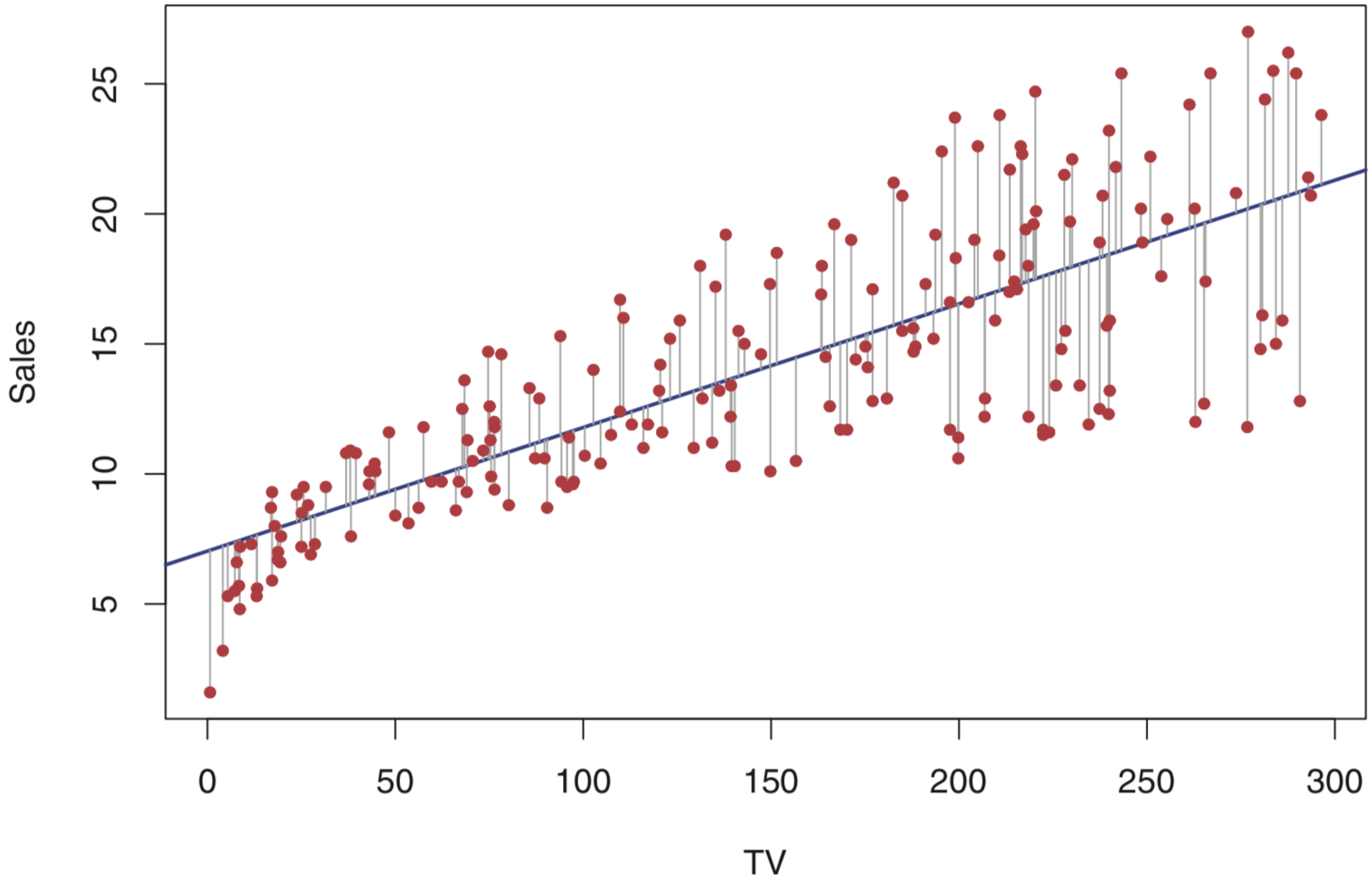
3) Is a linear model enough?

4) Can we predict y given a new x?

Regression Example



Example: predict sales from TV advertising budget



ISL: Figure 3.1

Cost Function: sum of squared errors