Machine learning algorithms

Generalization & bias-variance tradeoff

2020–10-06

CSCI 471 / 571, Fall 2020
Kameron Decker Harris

Recommended talk!

- Nathan Kutz, 3 p.m. on Thursday
- Western Washington Data-driven Discovery Seminar Series
- *Machine Learning for Science: Data-Driven Discovery Methods for Governing equations, Coordinates and Sensors*

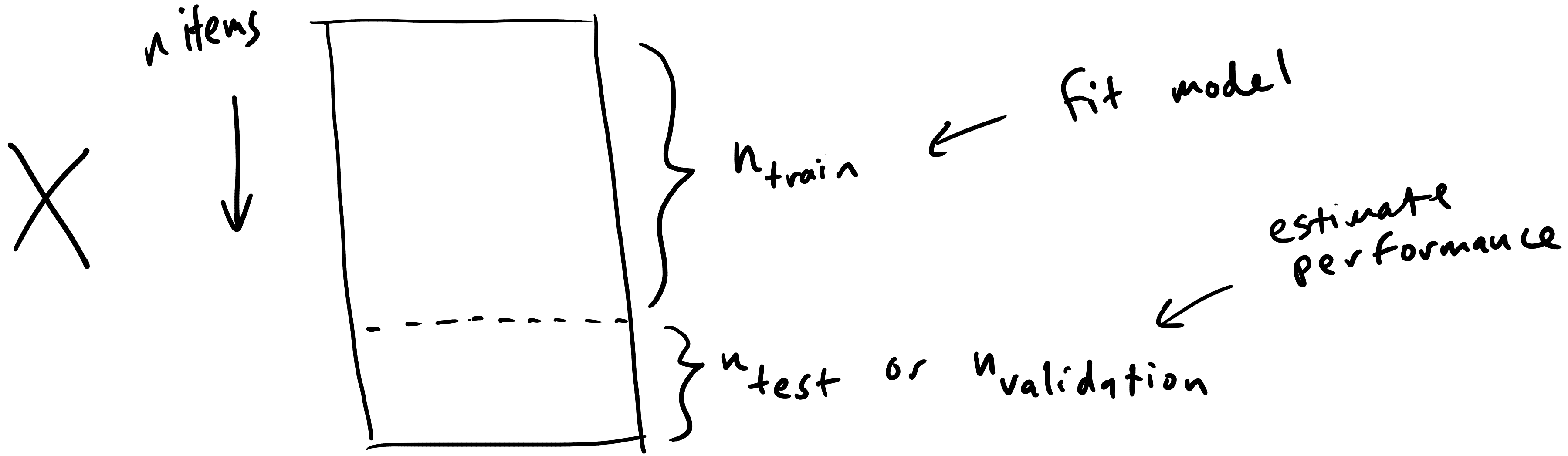
Homework questions?

Linear algebra review:
will be uploaded soon

Numpy help: tomorrow 4 pm

Generalization

- Making predictions on unseen data
- Can estimate with training/validation split



Comical consequences for errors

- “News broadcast triggers Amazon Alexa devices to purchase dollhouses”
- “Amazon Alexa starts a party -- and the neighbors call the cops”
- “Supposedly kid-friendly robot goes crazy and injures a young boy”

https://www.entrepreneur.com/slideshow/289621

Warning: potentially offensive examples ahead

Serious consequences for errors

Racism or discrimination

Google Photos mislabeled black people as “gorillas”
Yahoo Finance article

Tempe police

Car crashes

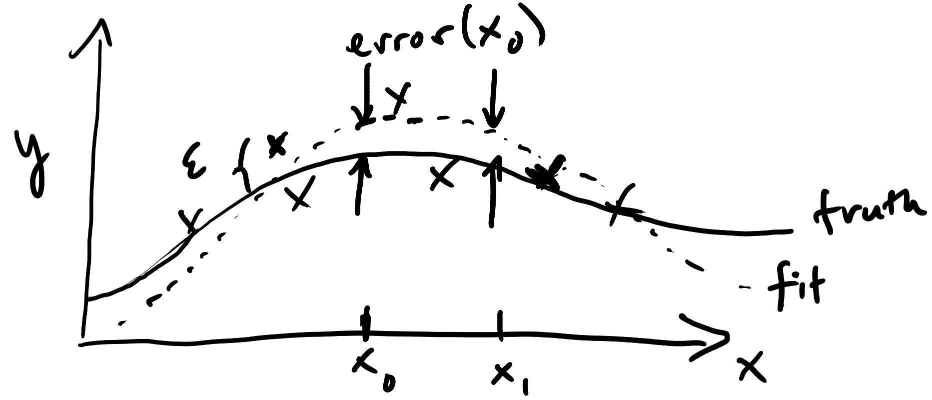
Uber self-driving car killed Elaine Herzberg
“Safety driver” charged but not Uber
Wired article

Rules for ML data

- Training data as close as possible to data used in application
  - avoid different train distributions from applied "distribution shift"
- Try to break your algorithm
  - testing for biases (race, gender, etc)
- Unbalanced classes require care
  - algorithms tend to fit to majority
  - balancing can be done

ISLR Sec 2.2 (James, Witten, ...)

(statistical) bias-variance tradeoff



$$y = \underbrace{f^*(\vec{x})}_{\text{true target function}} + \underbrace{\varepsilon}_{\text{noise}}$$

$\mathbb{E}[\varepsilon] = 0$

$\mathbb{E}[\varepsilon^2] = \sigma^2$

$$\hat{f} = \arg\min_{f \in \mathcal{F}} \sum_{i=1}^{n} (y_i - f(\vec{x}_i))^2 \leftarrow \text{squared error} = \text{training error}$$

$\mathcal{F}$ ← linear, polynomial, etc.

Goal: $\hat{f} \approx f^*$

$$\text{MSE}(\vec{x}) = \mathbb{E}\left[(\hat{f}(\vec{x}) - y)^2\right] = \underbrace{\sigma^2}_{\text{irreducible error}} + \text{Bias}^2(\vec{x}) + \text{Var}(\vec{x})$$

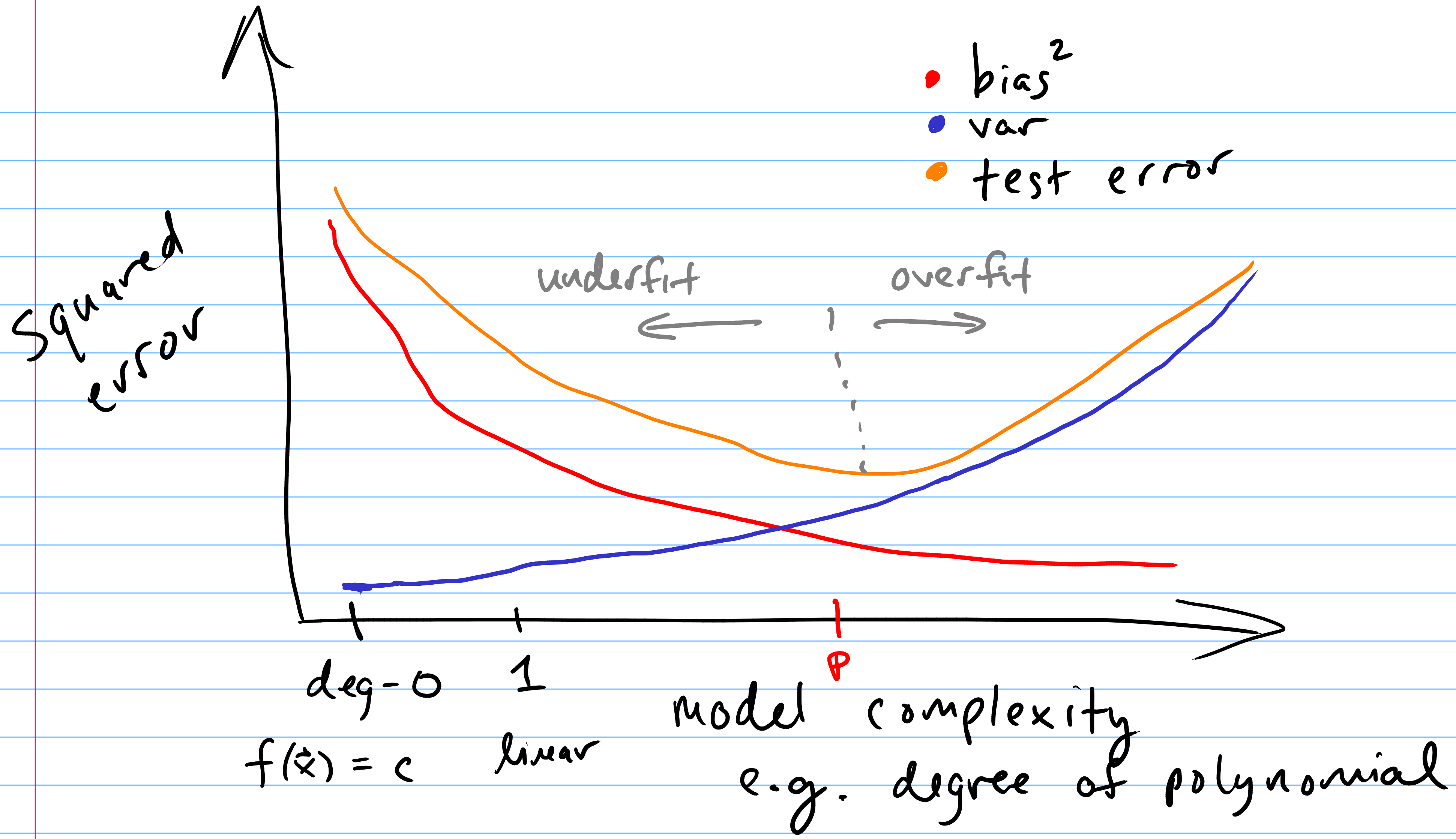
MSE: mean square error (testing)

$\mathbb{E}$: avg over all randomness

$\text{Bias} = \mathbb{E}[\hat{f}(\vec{x})] - f^*(\vec{x})$

$\mathbb{E}[\hat{f}(\vec{x})] \leftarrow \hat{f}_n$

$\text{Var}(x) = \mathbb{E}[\hat{f}(x) - \mathbb{E}[\hat{f}(x)]]^2$


bias^2
var
test error
squared error
underfit
overfit
deg-0
$f(\vec{x}) = c$
1
linear
p
model complexity
e.g. degree of polynomial