HUDK 4050: CORE METHODS IN EDM

11/20/18 5:43 PM

Qingying Zhou
Jiaying (Mandy) Li
Melissa Mock
Abdul Abad

In the news

K–12's Digital Transformation Is Giving Libraries a Modern Makeover

EdTech Focus On K-12

Facebook Launches Courses to Help Adult Learners Skill Up

EdSurge

TVs Built Into Gas Station Pumps Offer Information Bonanza For Advertisers

npr one

How Machine Learning Is Helping Data Centers Chill Out

Analytics India MAGAZINE

TECHSPOT

Data centers are using two percent of the world's energy

Events

Event	Date	URL
AI & Society -- Update on the current state of AI	6:00pm November 20	https://www.meetup.com/NYHAIS/?gj=wcs1_g&rv=wcs1_g&_xtd=gatlbWFpbF9jbGlja9oAJGIyZjRmNDM5LWFkNjMtNDc2Yi05YWEzLWY3ZWQwNTVmYTUxZA&_cookie-check=pG-mvPnB-VvxZan1
Business Analytics Poster Session	6:00pm November 26	datascience.columbia.edu
Computing for the Endless Frontier	2:30pm November 27	https://events.columbia.edu/cal/event/eventView.do?b=de&calPath=%2Fpublic%2Fcals%2FMainCal&guid=CAL-00bb9e25-666e323b3-01666-e4d9e20e-00002992events@columbia.edu&recurrenceId=
Cross-device User Clustering at Adobe	5:30pm November 29	https://events.columbia.edu/cal/event/eventView.do?b=de&calPath=%2Fpublic%2Fcals%2FMainCal&guid=CAL-00bb9e28-655b8cee-0165-5dd5c72b-00001287events@columbia.edu&recurrenceId=
New York Tech Zine Fair	12:00pm December 1	http://techzinefair.org/
New President Investiture	9:30am December 7	https://listserv.tc.columbia.edu/t/2465234/33578280/12169/5/
Machine Learning Innovation Summit	December 12-13	https://www.theinnovationenterprise.com/summits/machine-learning-innovation-summit-new-york-2018
Columbia Data Science Day	April 3, 2019	https://www.eventbrite.com/e/data-science-day-columbia-university-2019-tickets-49454358317

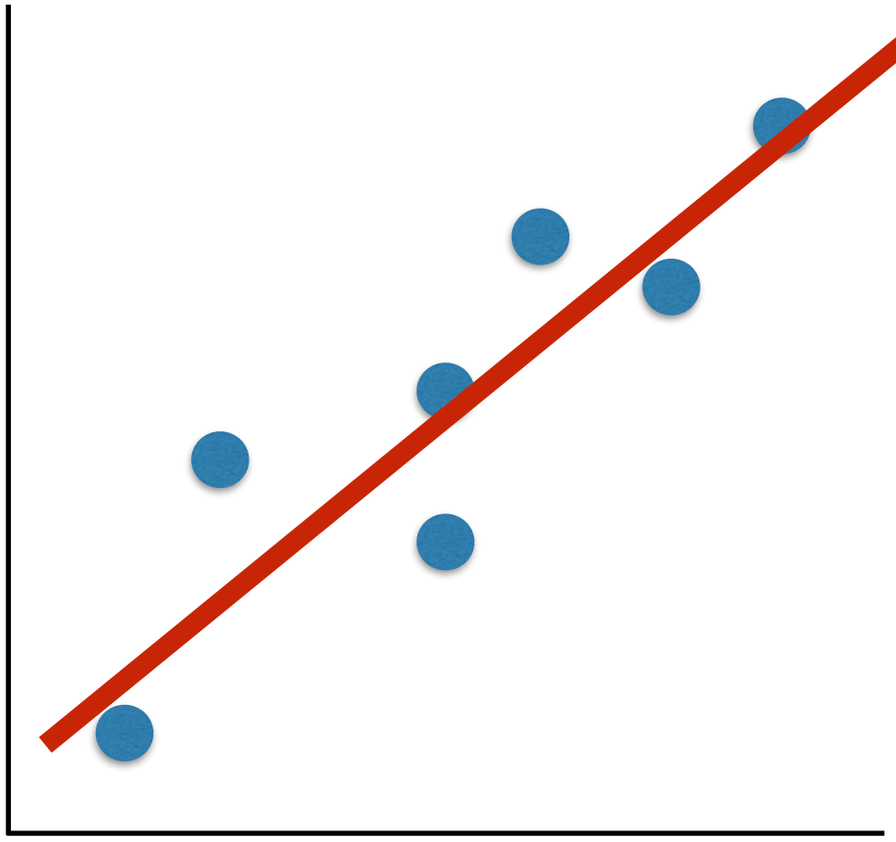
Chapter 6

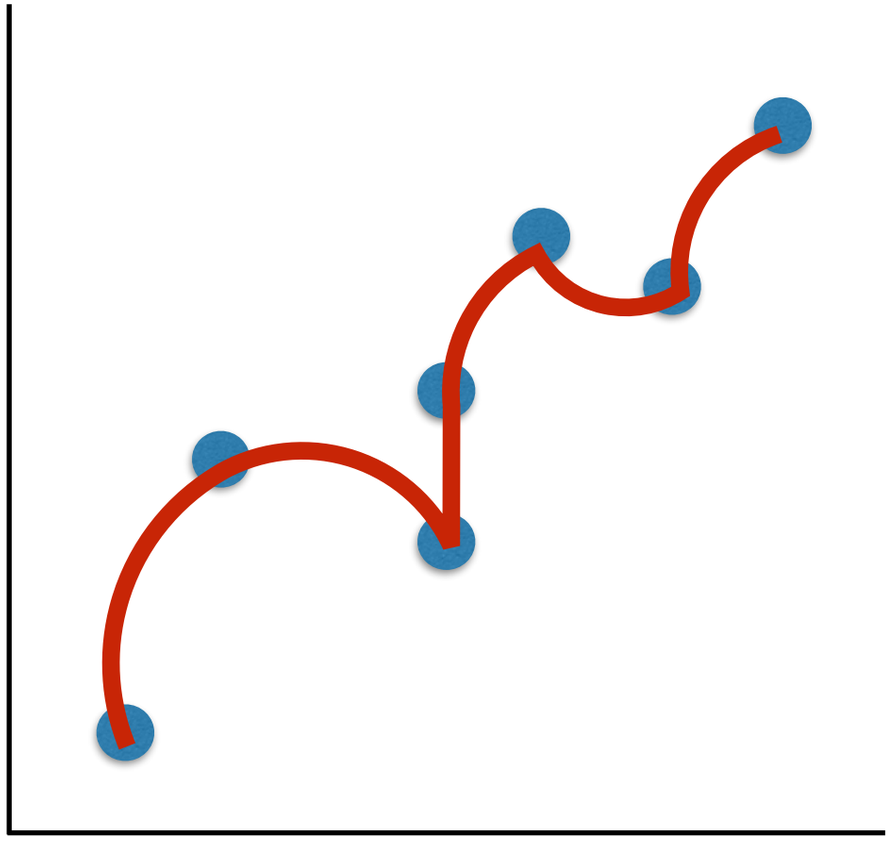
Handbook of Learning Analytics
First Edition

Going Beyond Better Data Prediction to Create Explanatory Models of Educational Data

Ran Liu & Kenneth R. Koedinger

Cross Validation

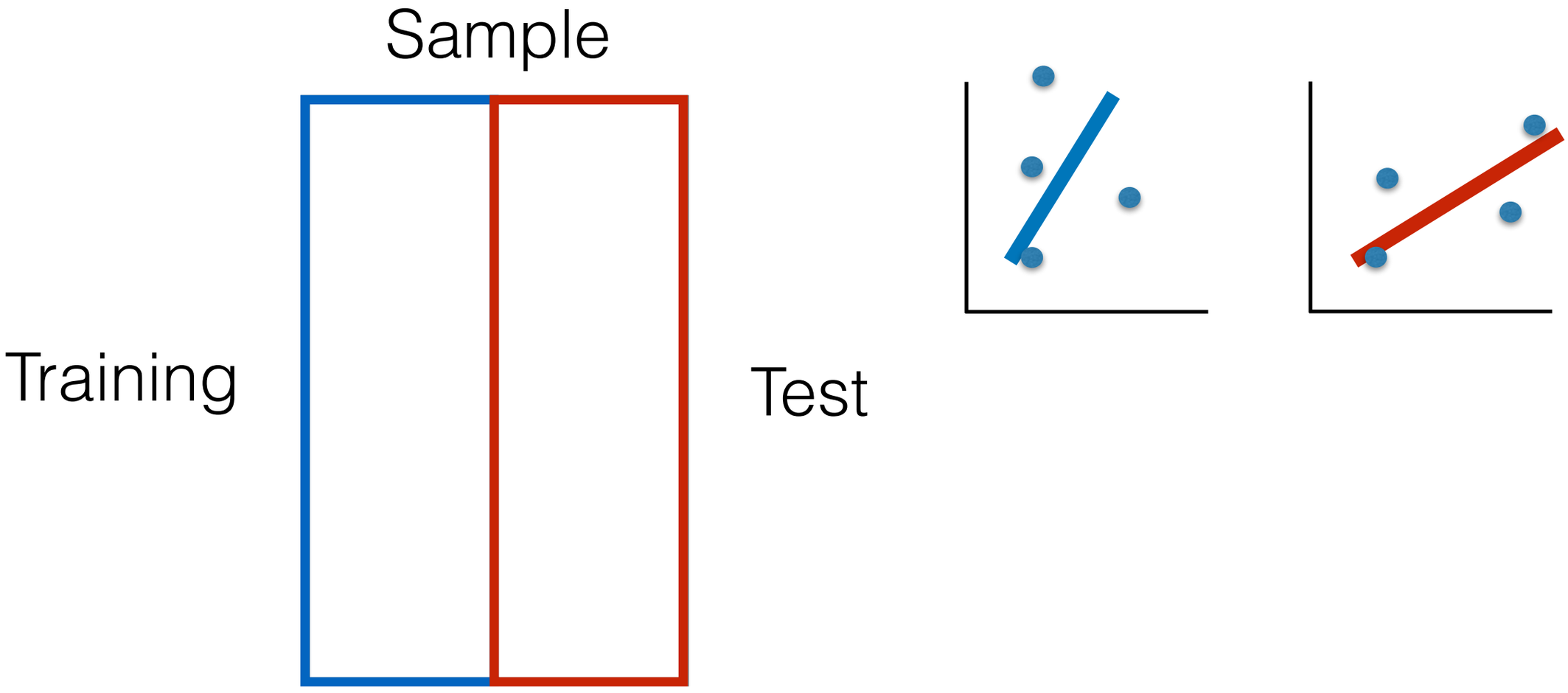




Cross Validation

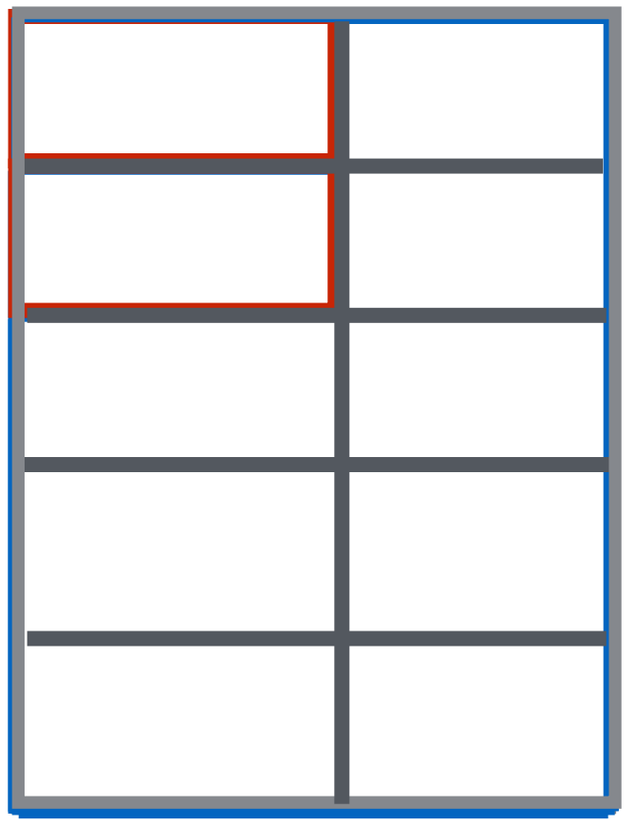
- Estimate how accurately a predictive model will perform in practice
- Give an insight on how the model will generalize to an independent dataset

Hold-out Validation



Problem: very dependent on which data are in each group

K-Fold Cross Validation



	M1 Error	M2 Error
	5	2
	4	2
	3	1
	5	4
	4	2
	4.2	2.2

Calculate how accurate we are in each “fold” and average the answer

Activity

Which make better pets? Dogs or cats?

Model	Prediction	Sample Error	Wild Sample
Random	dogs		
Majority Class	dogs		
Hold out Majority Class	cats	100%	
X-validated Majority Class	dogs, cats, cats, dogs, cats	1, 0, 1, 0, 1, 1	

RPART

```
      CP nsplit rel error xerror     xstd
1 0.052      0     1.000  1.072 0.035325
2 0.012      1     0.948  0.992 0.036941
3 0.010      2     0.936  1.004 0.036733
```

Assignment 6

core-methods-in-edm/assignment6