Q&A SESSION 1

Outline

- Week 1
  - Overview
  - Bias_variance
    (question from "Discussion forum")
  - K=1 in K-NN?
    (question from "Points to Think About")

WEEK 1

WHAT IS WEEK 1 ABOUT?

Introductions

Syllabus

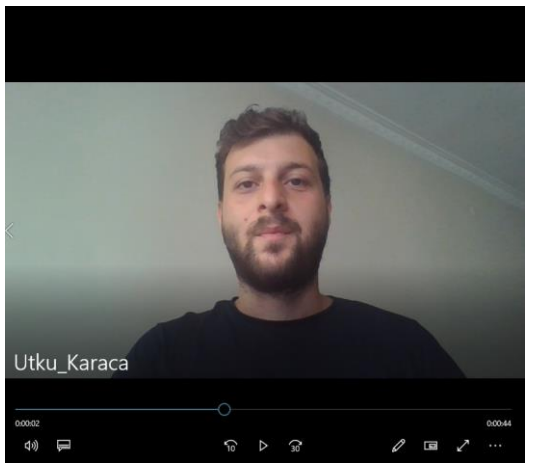
Python!

Lecture Notes

INTRODUCTIONS

Ilker Birbil

Utku Karaca

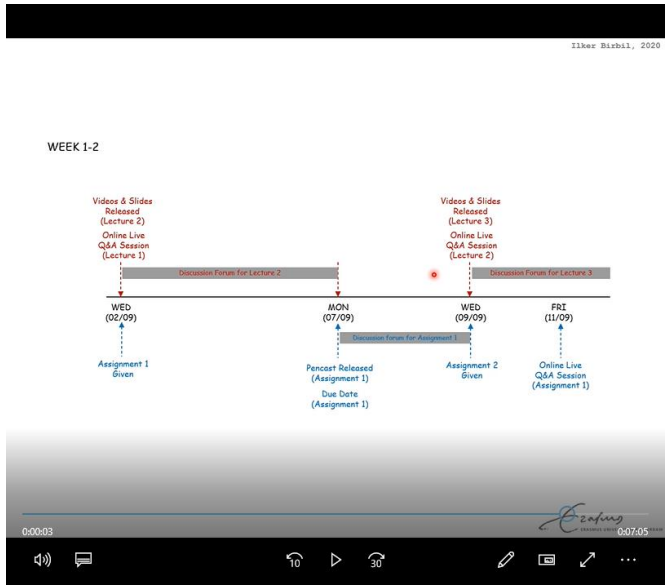


SYLLABUS

Course Plan

Week 1		Week 2	
Lecture		Lecture	Tutorial
Introduction		Regularization	Assignment 1
Assignment 1		Assignment 2	
Week 3		**Week 4**	
Lecture	Tutorial	Lecture	Tutorial
Trees, Forests & Boosting	Assignment 2	Unsupervised Learning	Assignment 3 (20%)
Assignment 3		Assignment 4	
Week 5		**Week 6**	
Lecture	Tutorial	Lecture	Tutorial
Support Vector Machines	Assignment 4	Neural Networks (Deep Learning)	Assignment 5 (20%)
Assignment 5		Assignment 6	
Week 7		**Week 8**	
Lecture	Tutorial		
Reinforcement Learning	Assignment 6	**EXAM (60%)**	

Detailed PDF file



Video description

Python

- Install

ANACONDA. ✓

- IDEs

SPYDER
The Scientific Python Development Environment ✓

Google colab + jupyter ✓

Visual Studio Code ✓

PyCharm ✓

THE TOP 10 LANGUAGES FOR MACHINE LEARNING HOSTED ON GITHUB

By Nick Heath

TechRepublic.

January 25, 2019

1. PYTHON

Highly rated machine-learning repositories

- sci-kit learn: Popular library for data mining and data analysis learning algorithms
- Machine Learning From Scratch: Bare bones but accessible in and algorithms
- ChatterBot: A machine learning, conversational dialog engine

OVERVIEW

Supervised learning

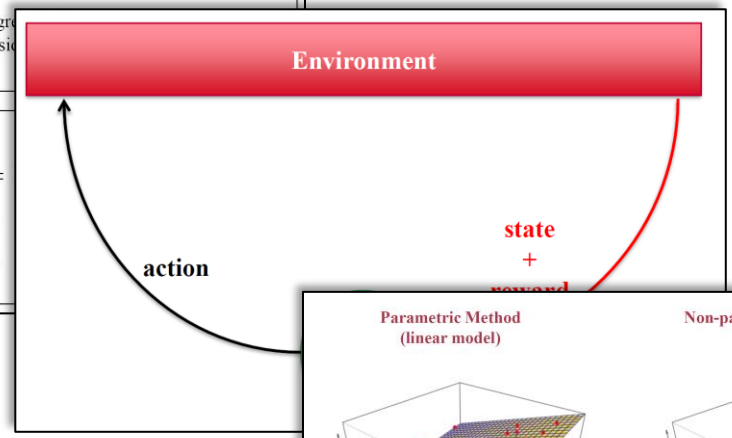
$$\{(x_i, y_i) : i = 1, \ldots, n\}$$

Spam filtering (classification)
Medical screening (classification/regr
Shopping amount prediction (regressi

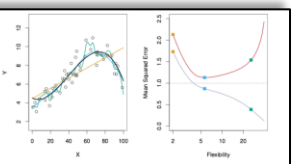
Unsupervised learning

$$\{x_i : i =$$

Customer clustering (segmentation)
Fraud detection (clustering)
Dimension reduction (compression)



Expected **test** MSE:

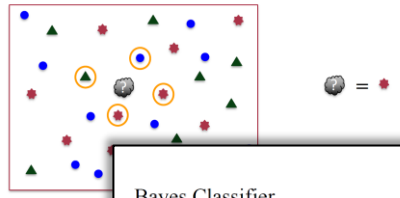


$$\mathbb{E}(y_0 - \hat{f}(x_0))^2 = \ldots$$
$$= \mathbb{E}((\hat{f}(x_0))^2) - \mathbb{E}(\hat{f}(x_0))^2 + (\mathbb{E}(\hat{f}(x_0)) - f(x_0))^2 + \mathrm{Var}(\epsilon)$$
$$= \mathrm{Var}($$

Roughly:

Model Complexity ↑

K – Nearest Neighbors



$$\mathbb{P}(Y = j|X = x_0)$$

Bayes Classifier

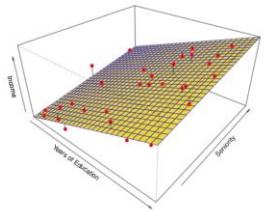

(source)

$$\hat{y}_0 = \arg\max_j \ \mathbb{P}(Y = j|X = x_0)$$

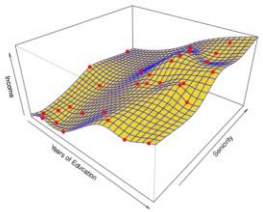
Bayes Error Rate

$$\mathbb{E}\big(1 - \max_j \ \mathbb{P}(Y = j|X)\big) = 1 - \mathbb{E}\big(\max_j \ \mathbb{P}(Y = j|X)\big)$$

Parametric Method (linear model)



Non-parametric Method (spline)

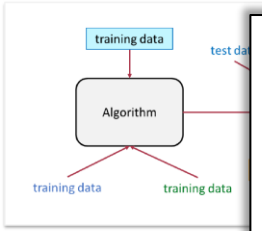


$$f(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

OVERVIEW

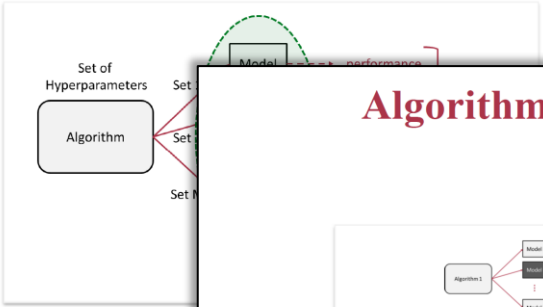
Generalization Performance

(algorithm and its parameters are fixed)

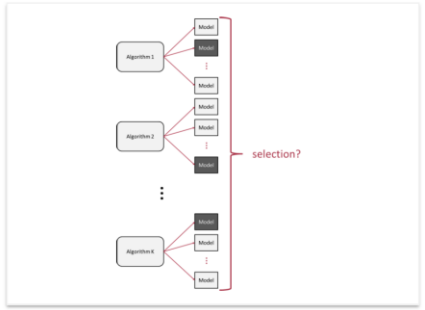


Model Selection

(algorithm is fixed, its 'best' parameters are sought)

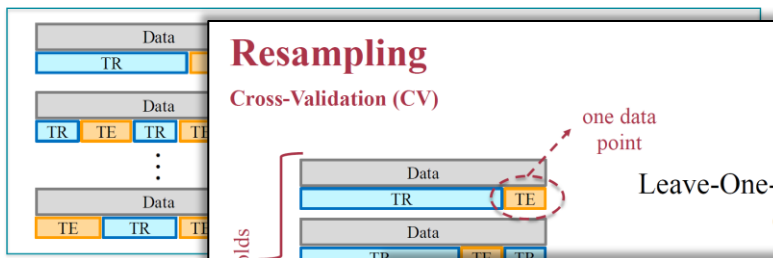


Algorithm Selection



Repeated Holdout

$i = 1, \ldots, k$



Resampling

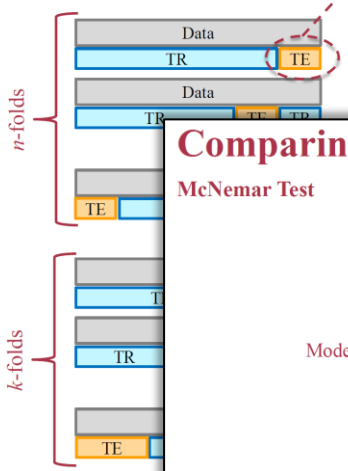
Cross-Validation (CV)

$\{(x_i, y_i) : 1, \ldots, n\}$

one data point

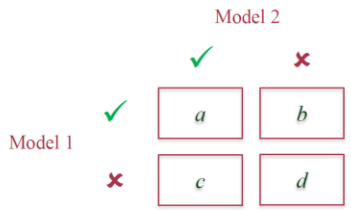
Leave-One-Out Cross-Validation (LOOCV)

n



Comparing Two Models

McNemar Test

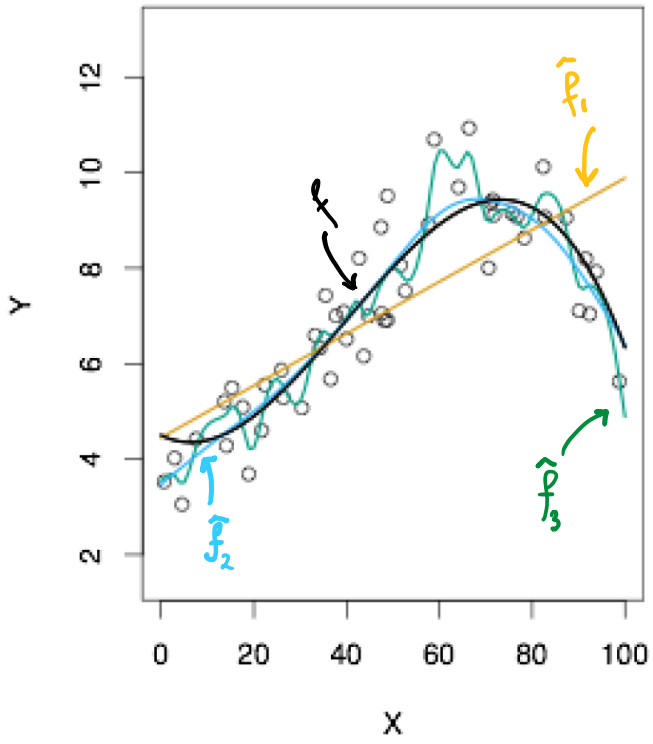


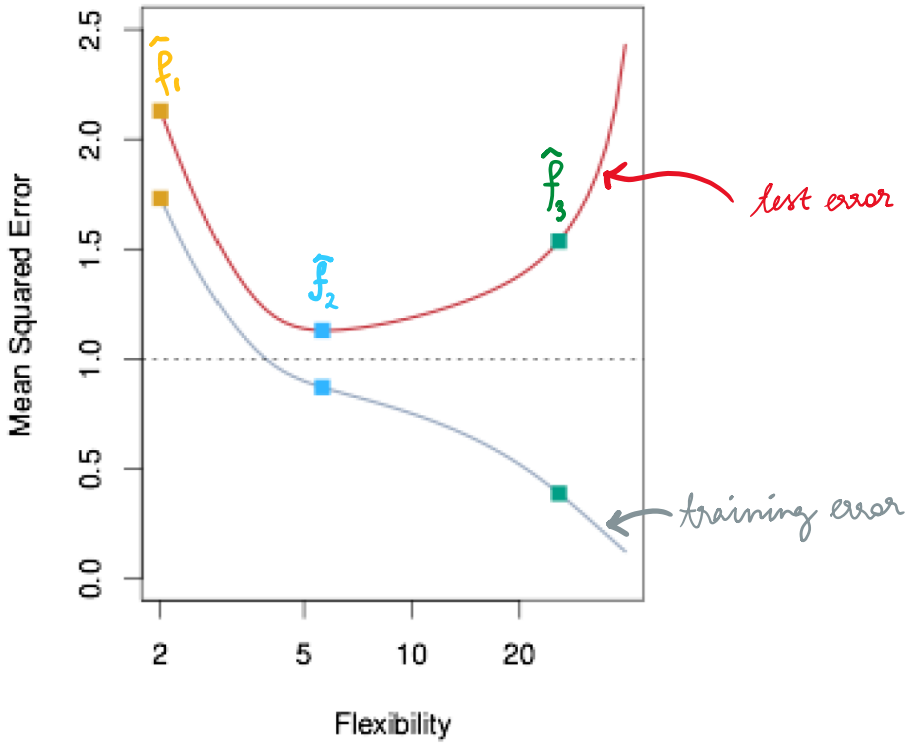
$a + b + c + d = n$

$H_0 : \mathbb{P}(b) = \mathbb{P}(c)$

Test statistic: $\chi^2 = \frac{(|b - c| - 1)^2}{b + c}$

BIAS - VARIANCE

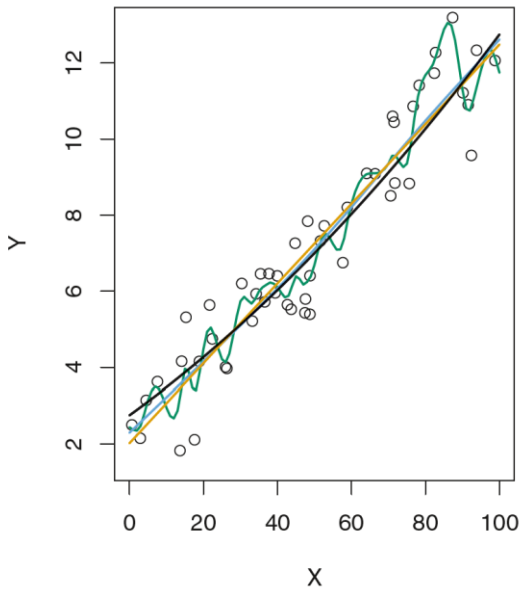


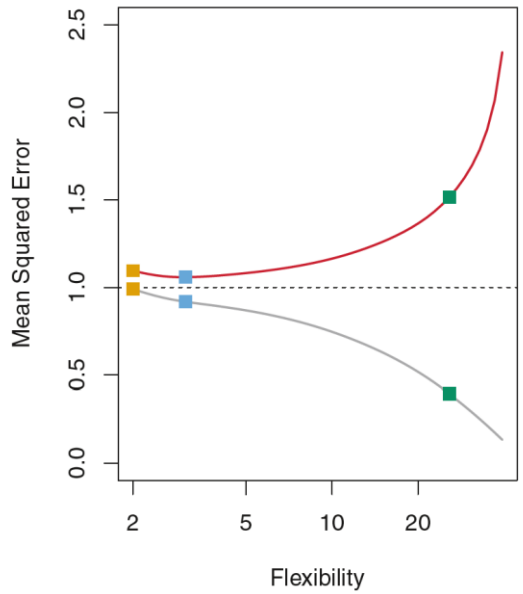


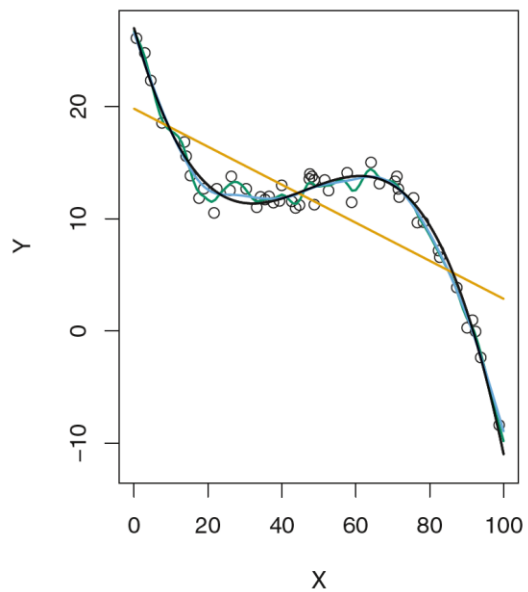
Roughly:

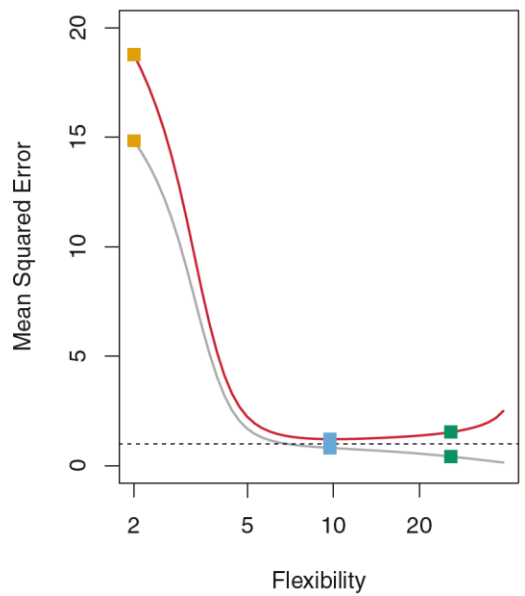
Model Complexity ↑	Variance ↑	Bias ↓

BIAS - VARIANCE

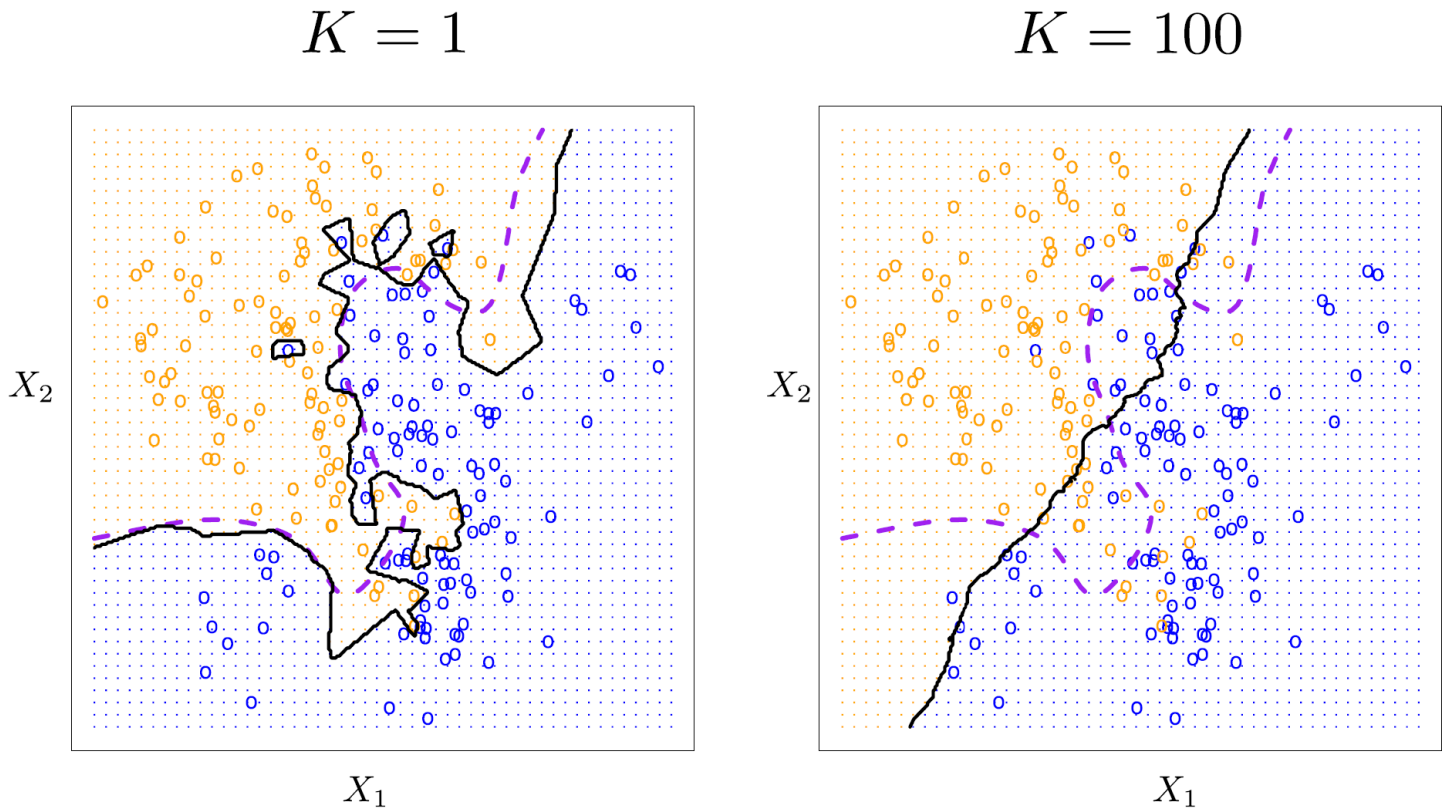








K=1 IN K-NN? Training Error = ?



Training Error = 0

TILL NEXT WEEK!

* Assignment 1 (Due date: 7 Sep.)

* Videos for lecture 2 (↓ 5pm)