391L Machine Learning

Instructor: Dana Ballard

Office Hours Tuesday 2-4pm
GDC 3.510

TA: Lijia Liu

Office Hrs: Thursday 2-4pm

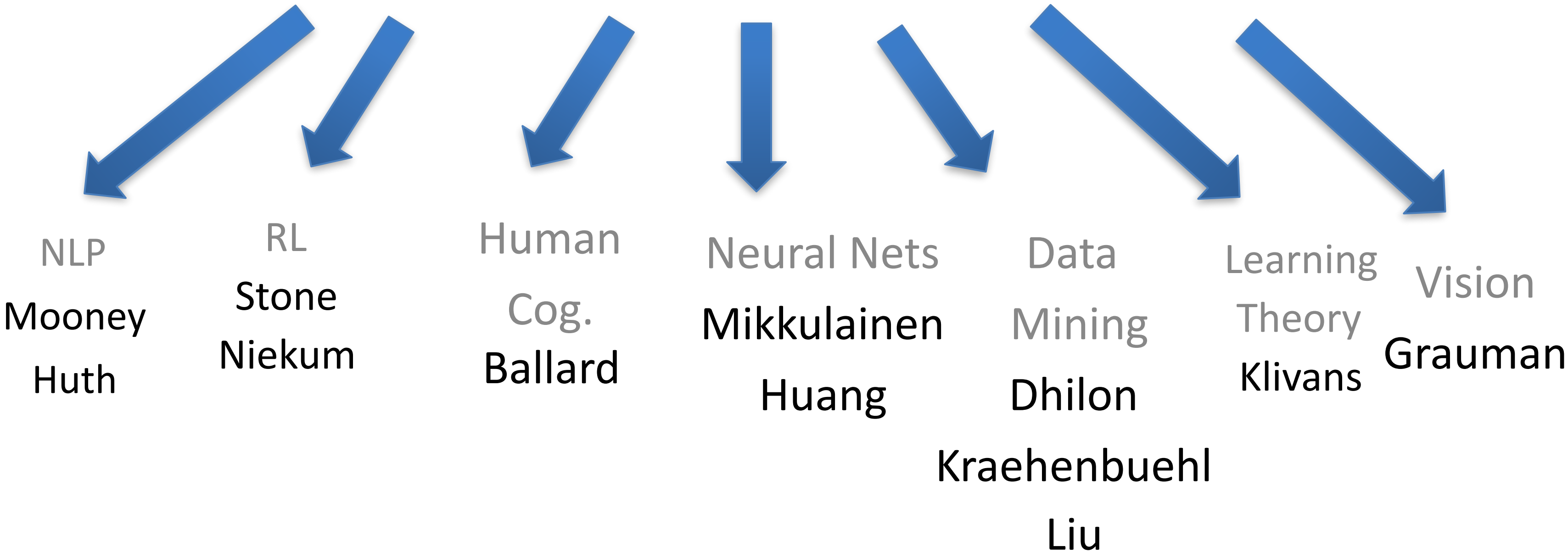
TA Station Desk 2

Required Text: Bishop
Pattern Recognition and Machine Learning

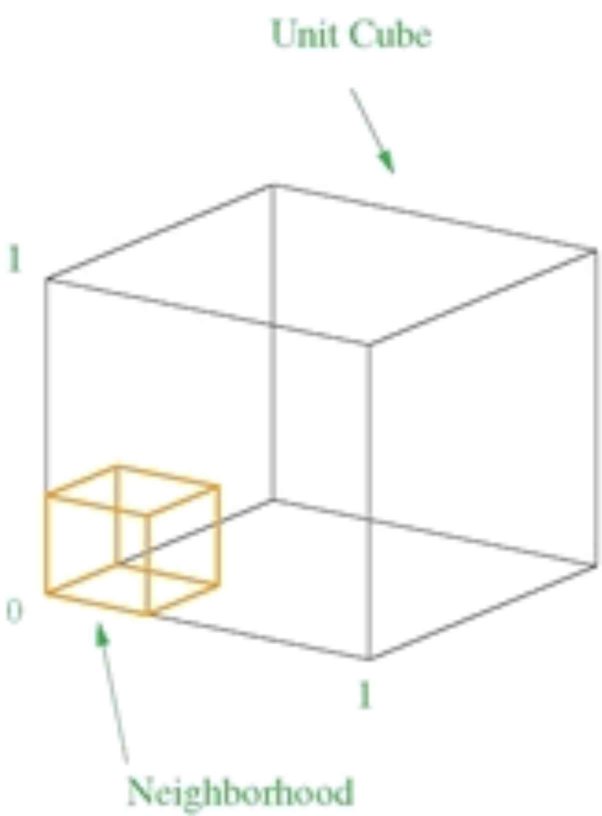
Recommended: Marsland
Machine Learning

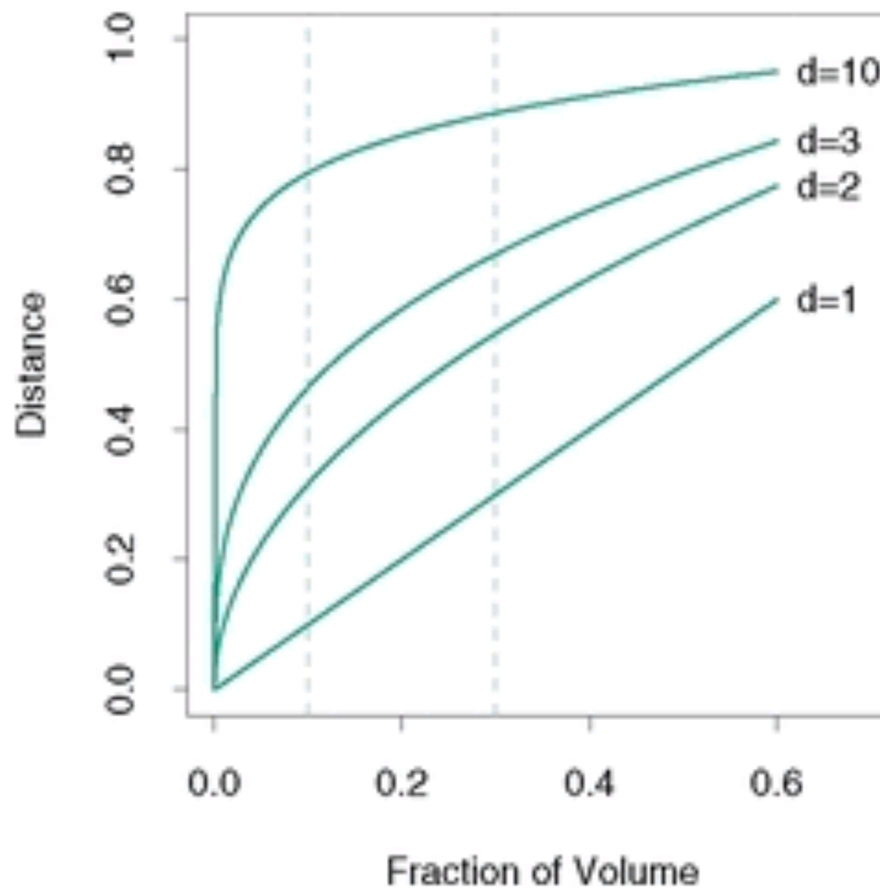
Six assignments: 60%
Two exams: 40%

391L Machine Learning



The Curse of Dimensionality





Volume of orange cube goes to … zero!

Random vectors $\mathbf{X}$

$$E(X) = \mu \qquad \text{Var}(X) = \sigma^2$$

Sum of andom vectors

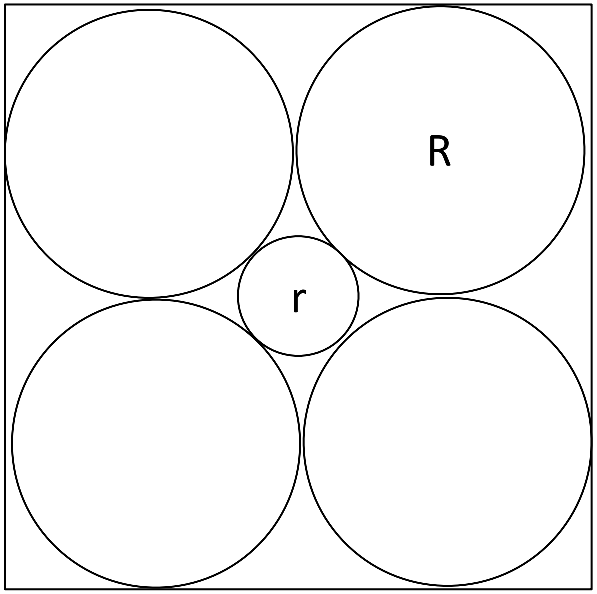
$$\mathrm{E}(\bar{X}) = \frac{1}{n}\mathrm{E}\left(X_1 + X_2 + \ldots + X_n\right) = \left(\frac{1}{n}\right)(n\mu) = \mu$$

$$\mathrm{Var}(\bar{X}) = \left(\frac{1}{n}\right)^2 \mathrm{Var}\left(X_1 + X_2 + \ldots + X_n\right) = \left(\frac{1}{n}\right)^2 (n\sigma^2) = \frac{\sigma^2}{n}$$

Distance between random vectors {-1,1} as *n* gets large?

High-dimensional spaces have counterintuitive properties

Two dimensional case



What happens to the ratio r/R as d -> infinity?

Main problems in Machine Laerning

Supervised

Classification   Use labeled training samples to build
a classifier that can label new samples

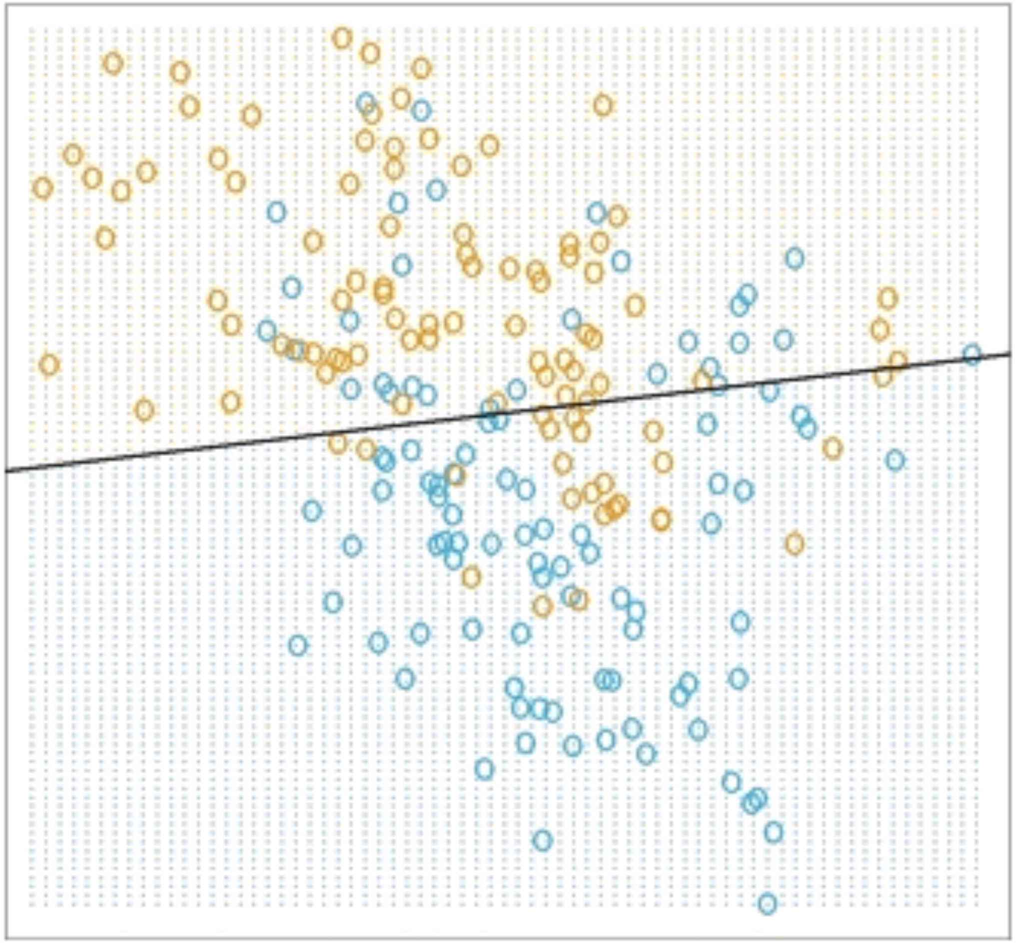
Regression   Use labeled training samples to fit
a curve that can predict new samples location

Unsupervised

Clustering
eg movie preferences

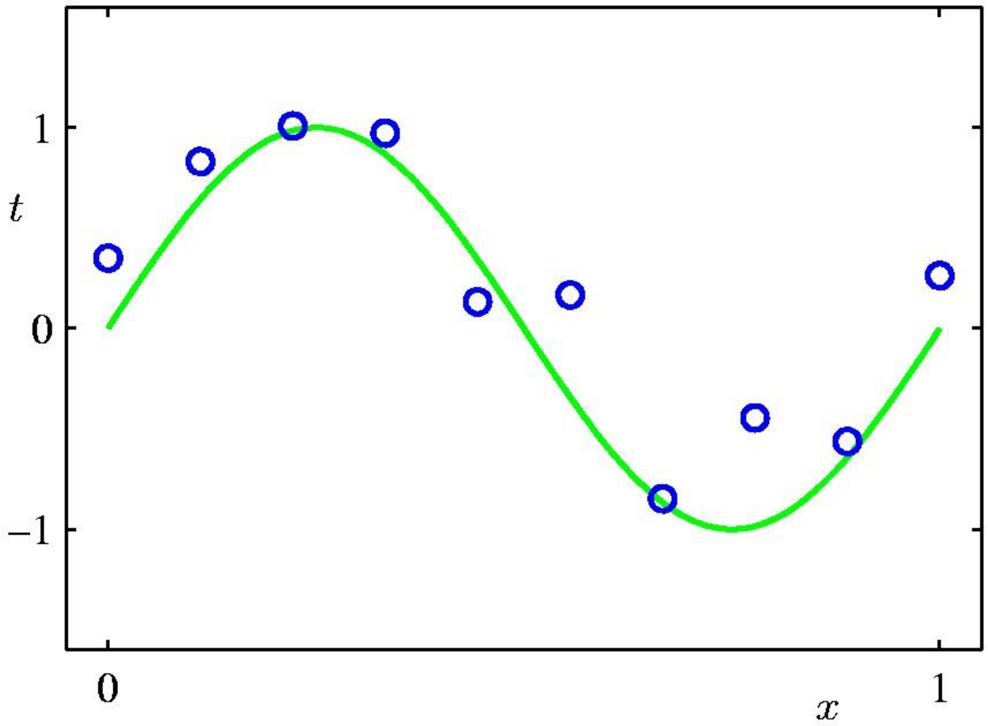
Classification:

Linear separation using least squares

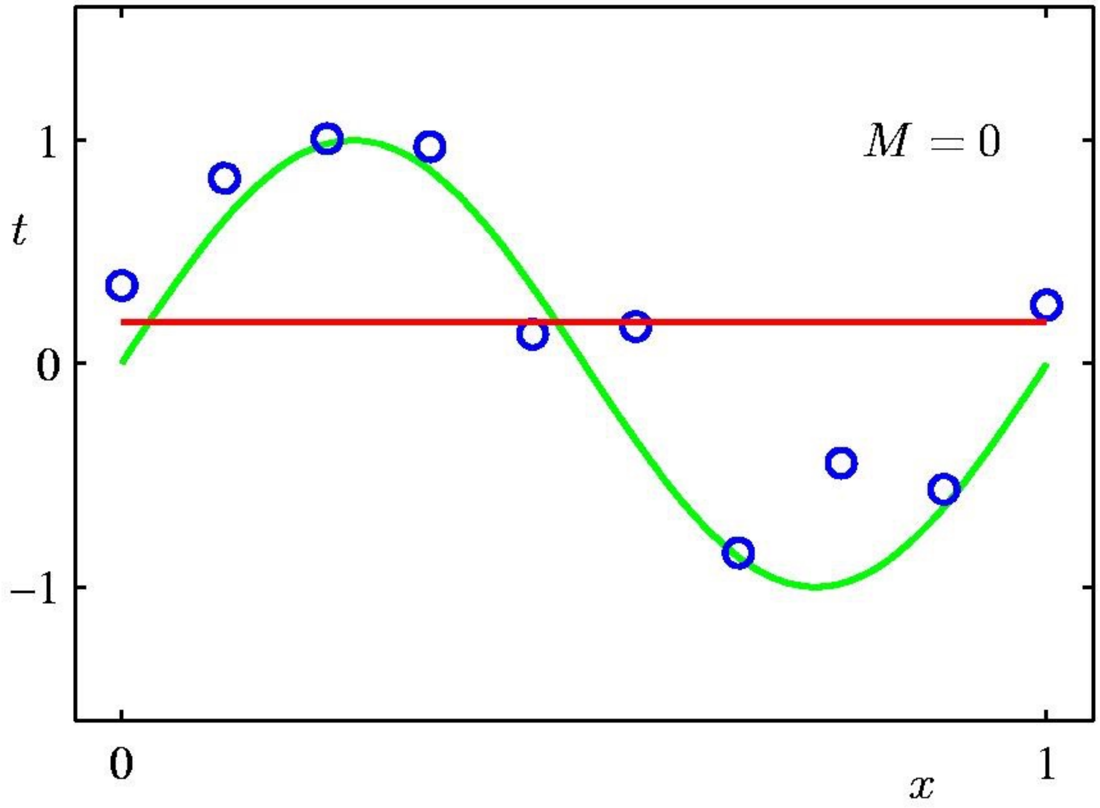


Regression

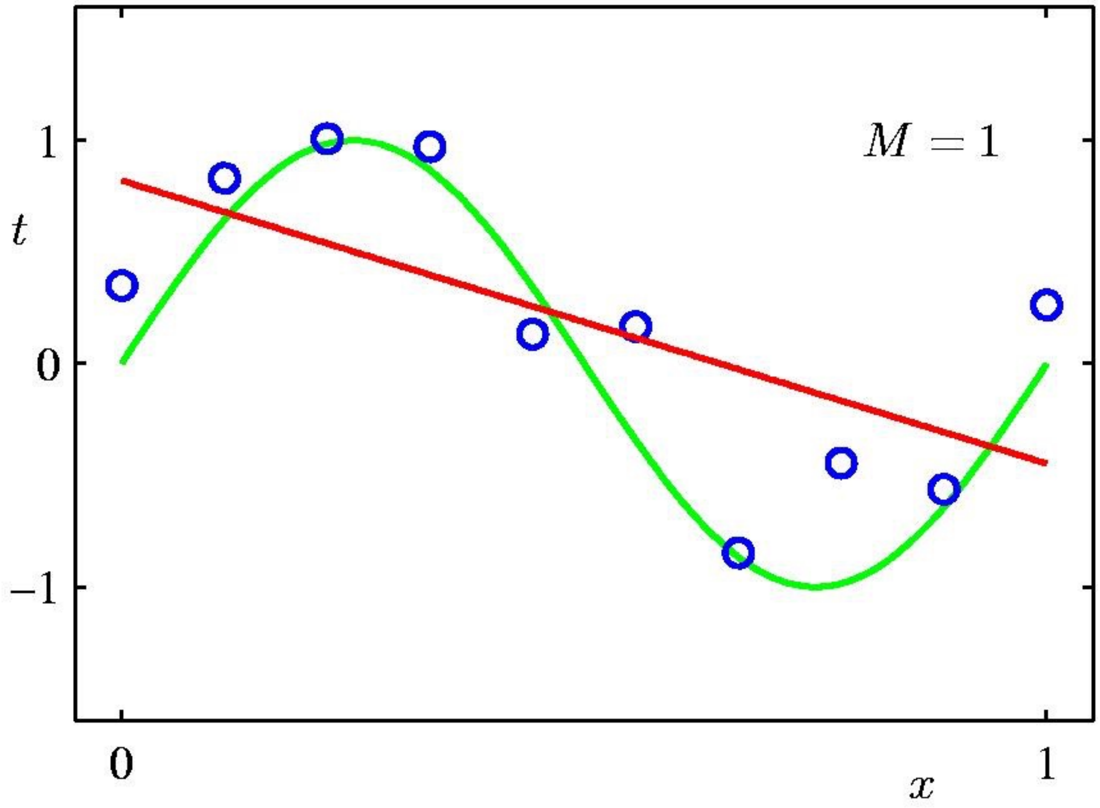
Polynomial Curve Fitting



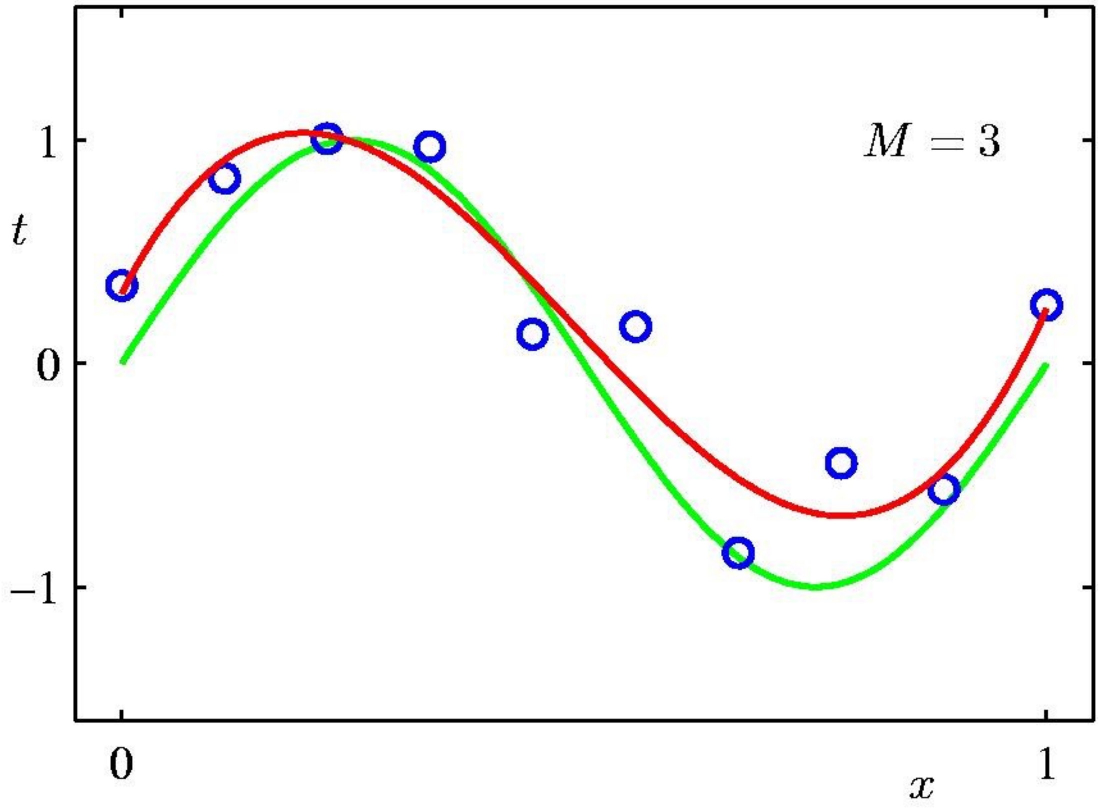
0^{th} Order Polynomial



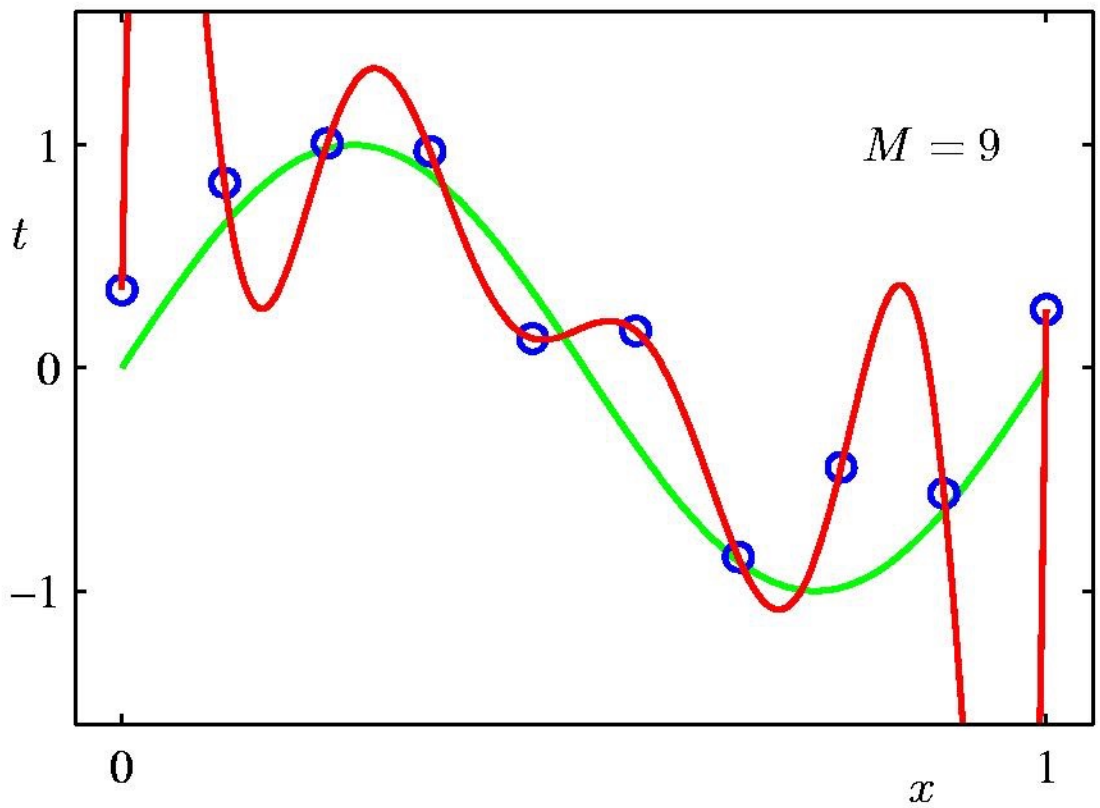
1st Order Polynomial



3rd Order Polynomial

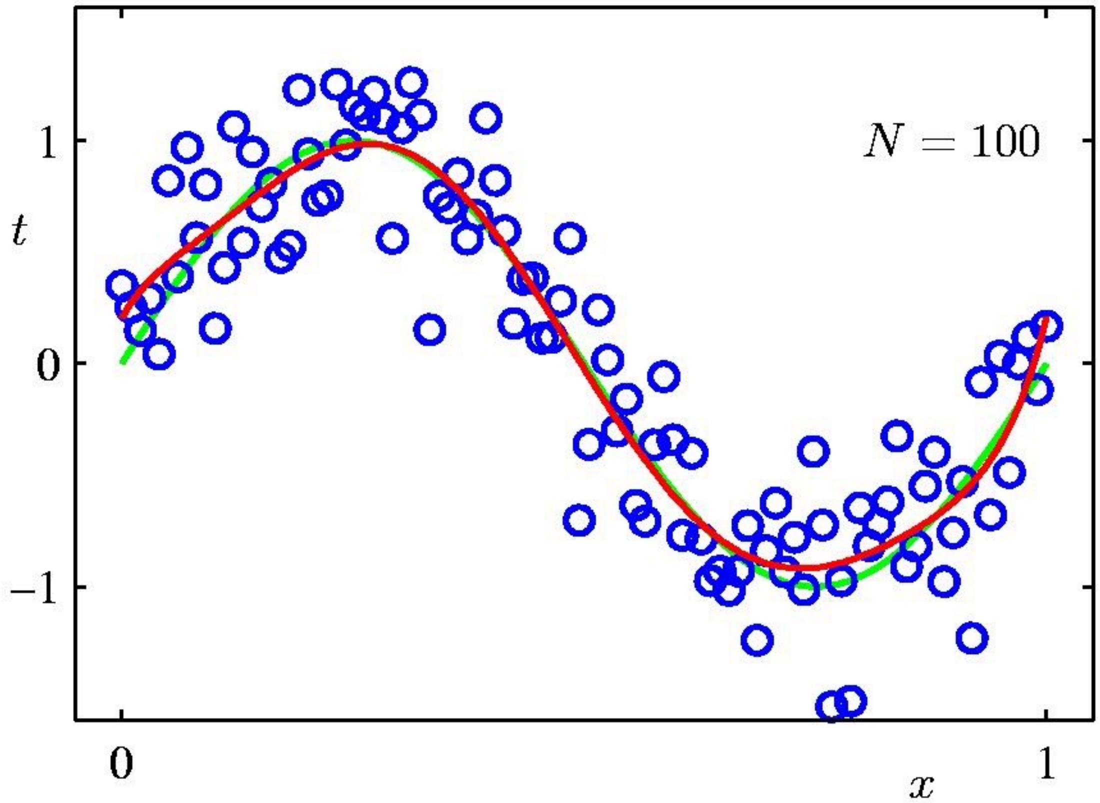


9th Order Polynomial



Data Set Size: $N = 100$

9th Order Polynomial



How much training data do we need?

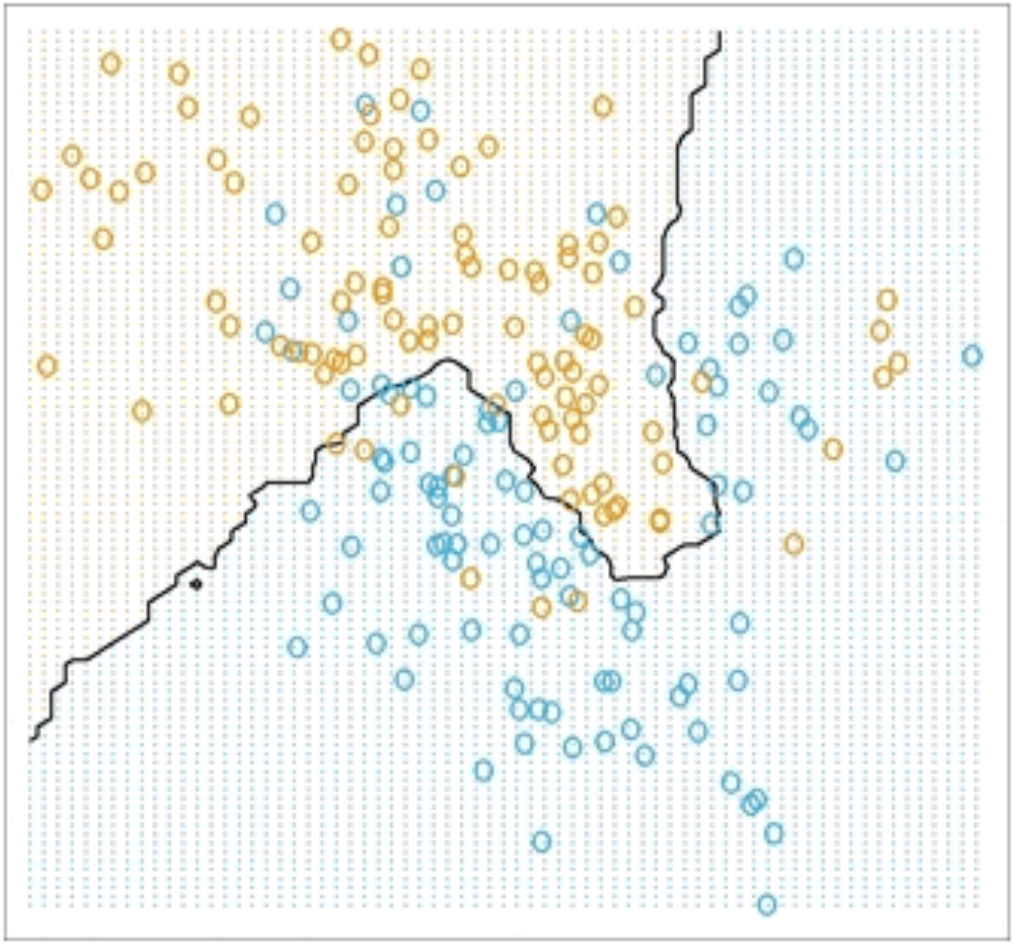
We want to be able to promise performance levels in the test phase

How much data do we need to guarantee a given performance level?

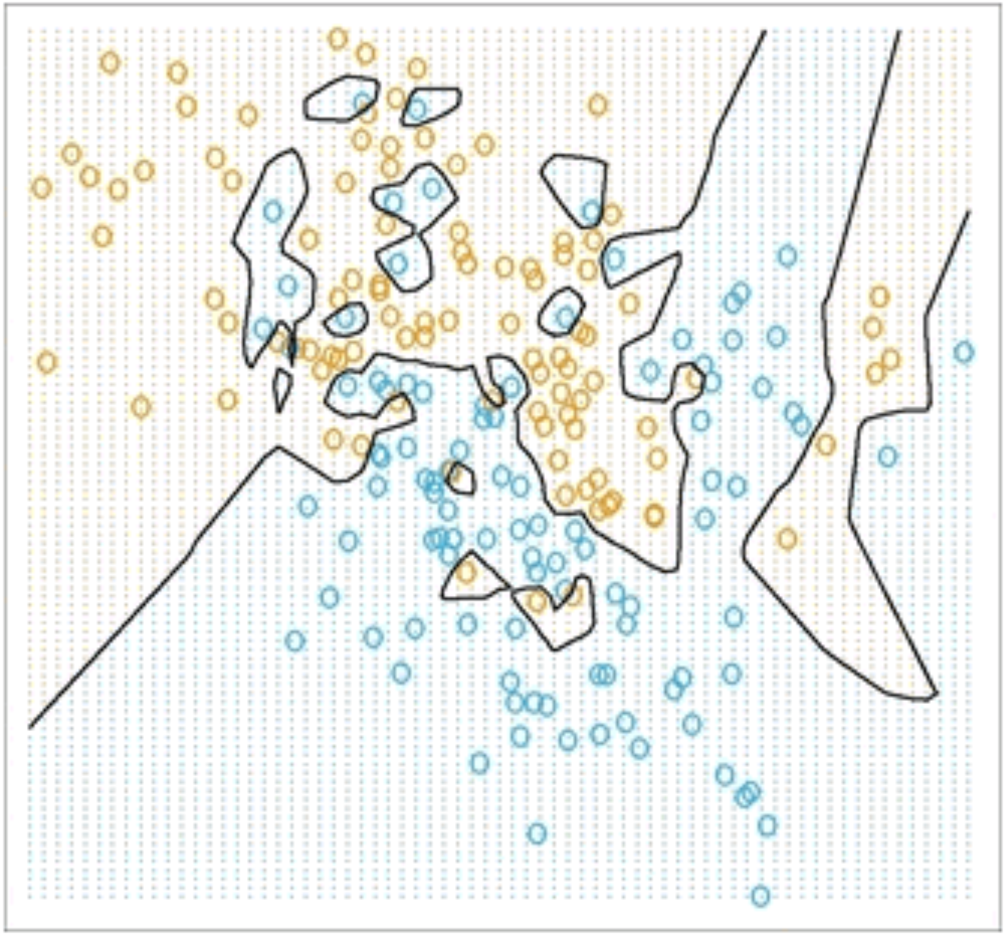
It turns out to be a function of the power of the function of the type of f

We use the terminology ‘machine’ to denote a class of f

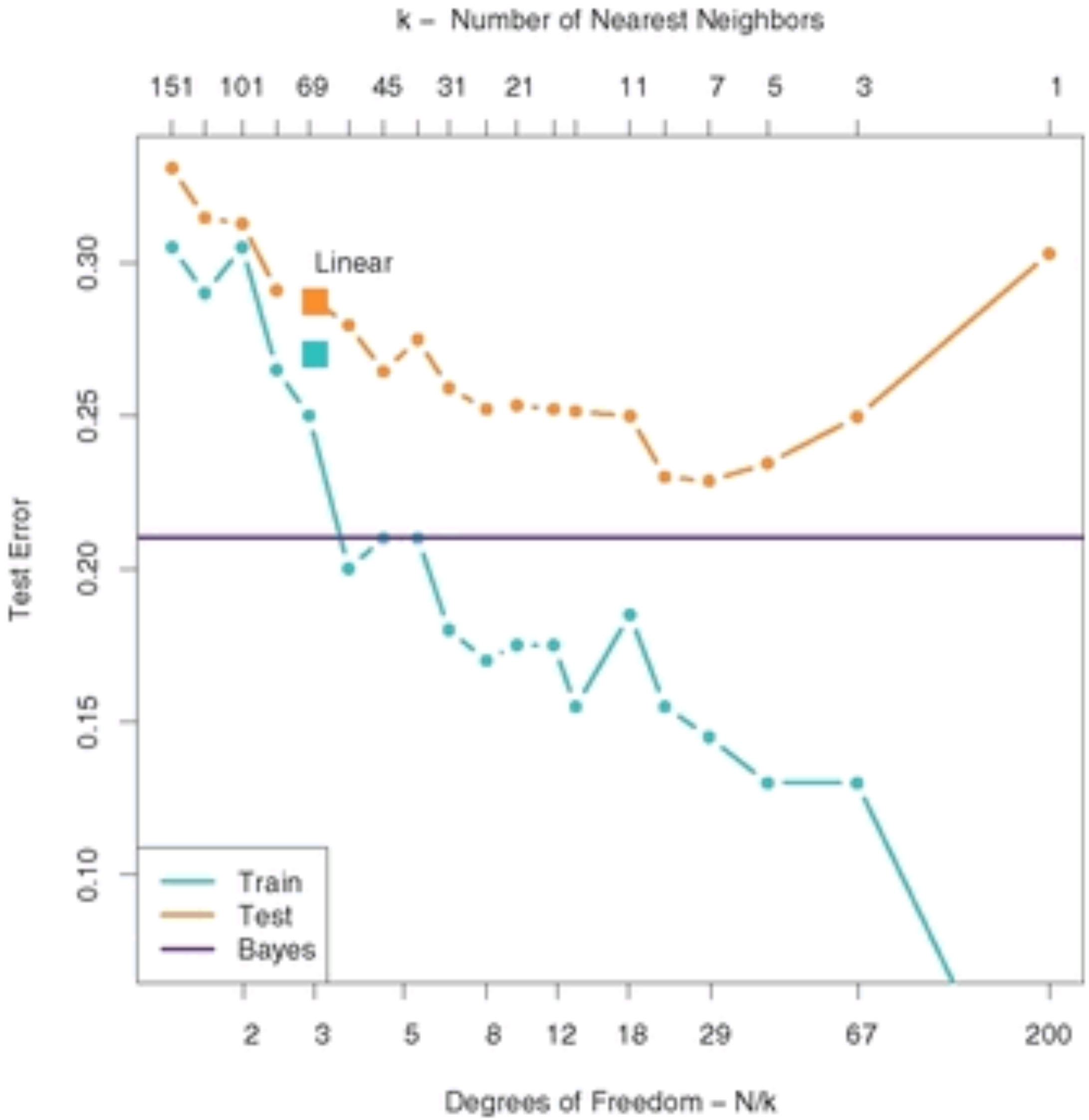
K Nearest Neighbour   k=15



NN k=1

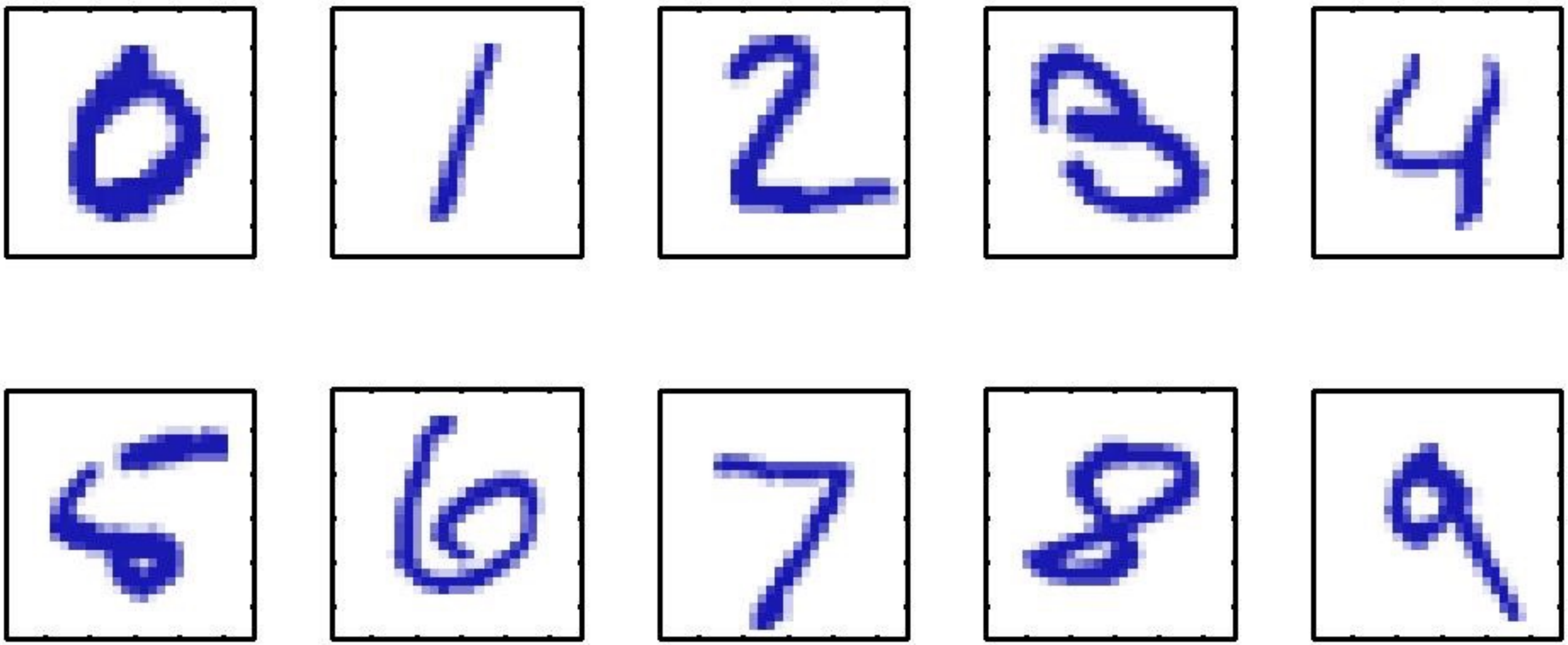


Linear vs. Nearest Neighbour for different k



Example 2 Handwritten Digit Recognition

28x28 = 784 pixels



9

Homework guide

The programming homework should have the format of a conference paper, with sections

Introduction

Methods

Results

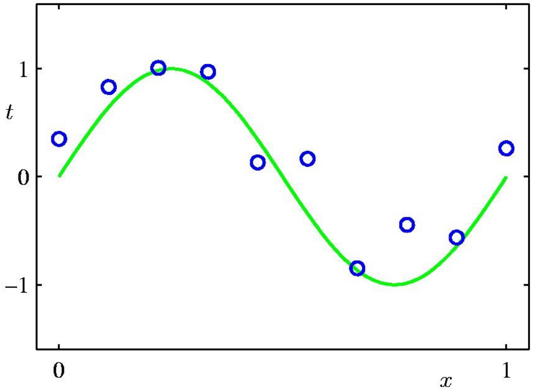
Conclusions

Link to code

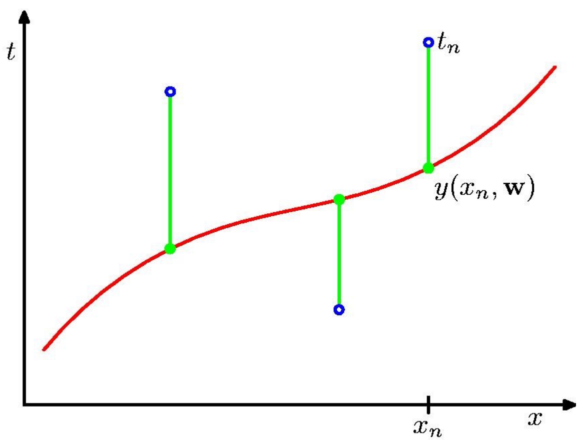
Week (Monday)	Monday	Wednesday	Bishop Chapter	Homework	Homework due (Friday)
Jan 20		Introduction			
Jan27	Linear Eqs	Eigenspaces	Notes	Eigendigits	
Feb 3	Probability Thy basics	Analytical Distrs	1 & 2		Eigendigits
feb 10	Information Thy mutual information KL divergence	ICA Ng derivation	Notes	ICA	
Feb 17	Sampling I analytical, Gauss importance	Sampling II MCMC, Gibbs	11		ICA
feb 24	Gaussian Process I	Gaussian Process II	Notes	Problem Set	
Mar 2	SVMs I basic eqns	SVMs II Learning params	7		Problem Set
Mar 9	Exam prep	Mid term exam		Gaussian Process	
Mar 16					
Mar 23	Hidden Markov Models	Reinforcement Learning	Notes		Gaussian Process
Mar 30	Reinforcement L I	Reinforcement L I	Notes	Reinforcement L	
Apr 6	Backpropagation	Convolution Nets	5, Notes		
Apri 13	Deep L	Thanksgiving	Notes, 6		Reinforcement L
Apr 20	Graphical Models I	Graphical Models 2	8	Deep L	
Apr 27	Graphical Models3	Learning Thy	Notes		Deep L
May 4	Exam prep	Final Exam			

Regression

Polynomial Curve Fitting

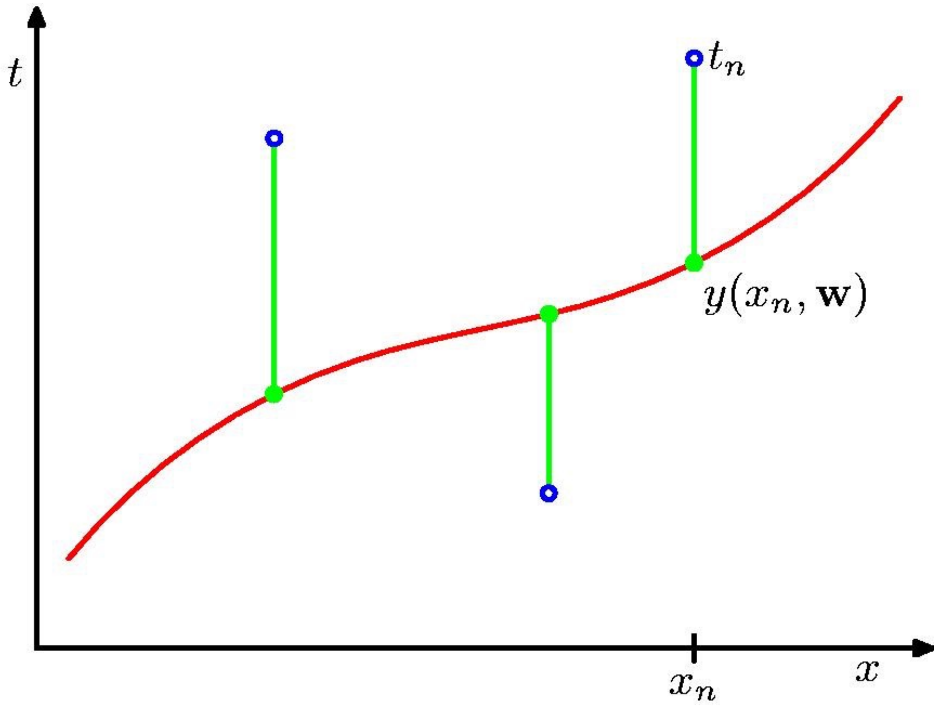


$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$



$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\}^2$$

Sum-of-Squares Error Function



$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2$$

Because **w** appears in the polynomial as a linear set of coefficients, E(**w**) has a straightforward solution.