DS-GA 3001.007 Introduction to Machine Learning

Lecture 2

Agenda

- Review
  - What are the different types of ML?
  - How to use ML for data science?
  - What are the components of ML?
- Lesson
  - Perceptron Algorithm
- Demo
  - Churn analysis

Reminders

- Section
  - Section 008 Thursdays 2:25-3:15 pm (60th 5th Avenue, Room 115)
  - Access to https://iml-f19.jupyter.hpc.nyu.edu
- Homework 1
  - Due September 18 11:59pm on Gradescope
- NYU Classes
  - Lessons
  - Demos
  - Labs

Types of Machine Learning

- Supervised Learning
  - Classification
  - Regression

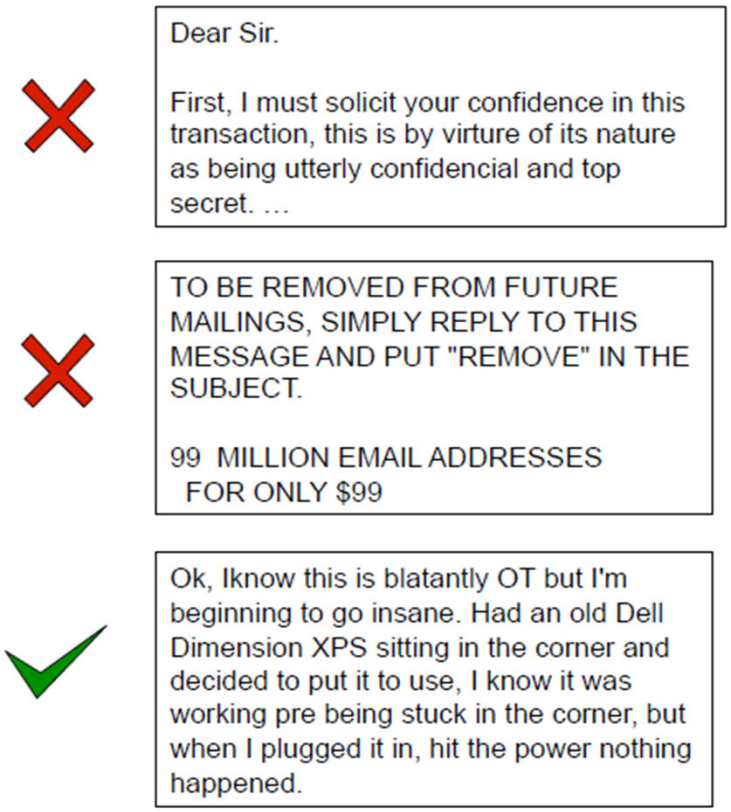
Types of Machine Learning

- Supervised Learning
  - Classification
  - Regression
- Unsupervised Learning
  - Clustering
  - Dimension Reduction

Types of Machine Learning

- Supervised Learning
  - Classification
  - Regression
- Unsupervised Learning
  - Clustering
  - Dimension Reduction
- Reinforcement Learning

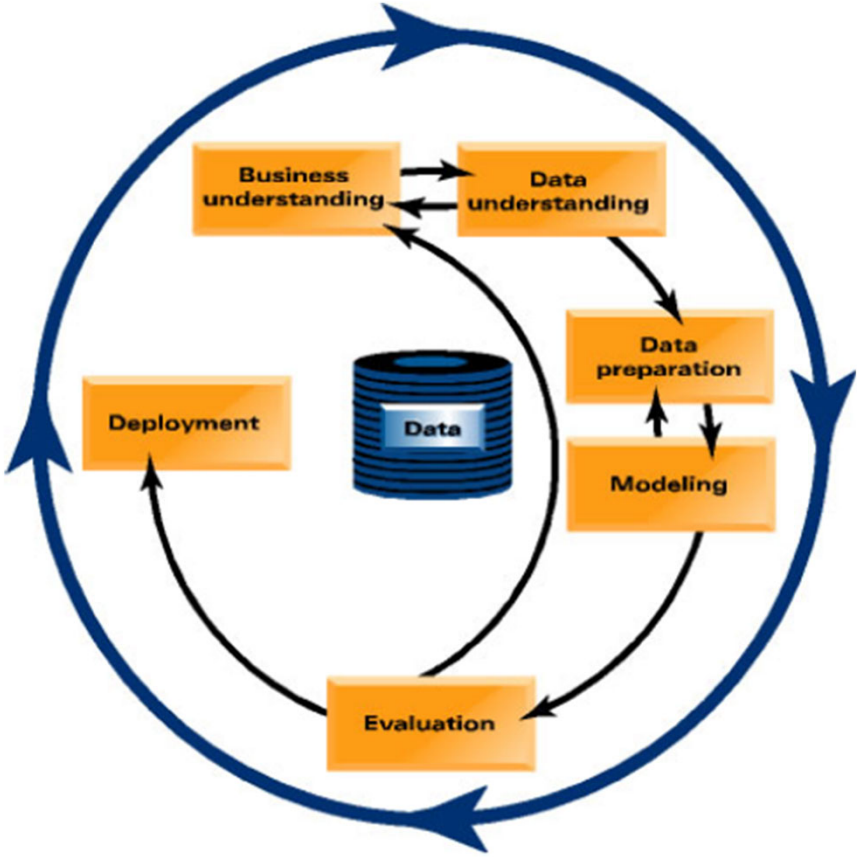
Types of Machine Learning



How to use machine learning for data science?

- Components of Data Science
  - Formulate a problem
  - Gather data
  - Explore data
  - Determine a model for prediction and inference
  - Evaluate findings

How to use machine learning for data science?



Components of Machine Learning Application

- Sampling data
  - Train Sample
  - Validation Sample
  - Testing Sample

Components of Machine Learning Application

- Sampling data
  - Train Sample
  - Validation Sample
  - Testing Sample
- Input/Output
  - Features
  - Labels

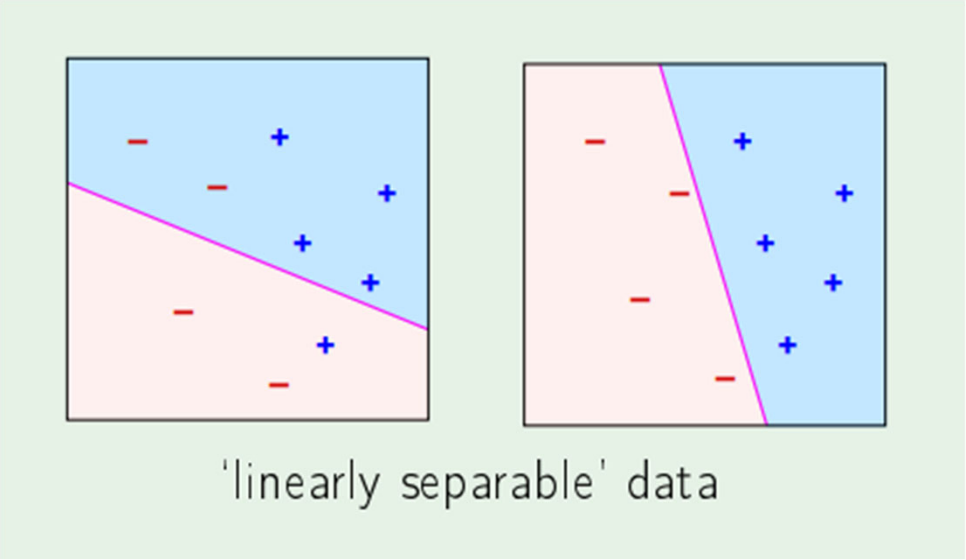
Components of Machine Learning Application

- Sampling data
  - Train Sample
  - Validation Sample
  - Testing Sample
- Input/Output
  - Features
  - Labels
- Fitting a model
  - Hypotheses
  - Hyperparameters

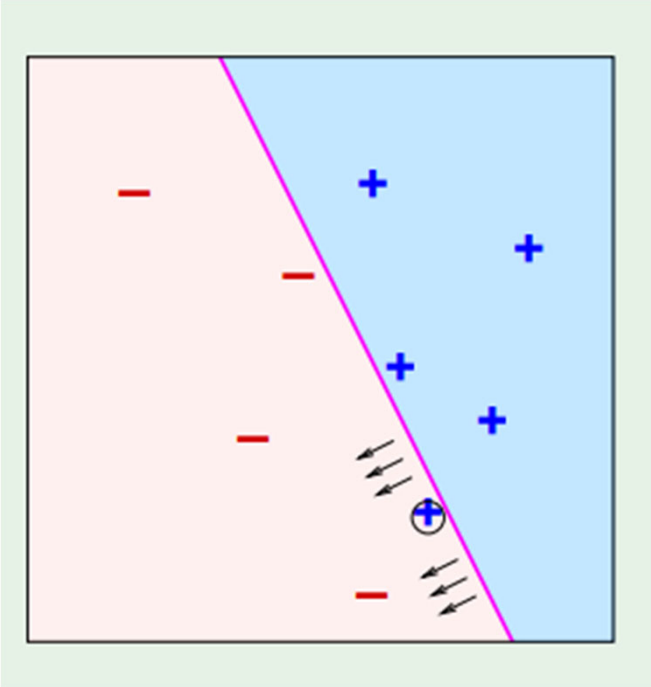
Components of Machine Learning Application

- Sampling data
  - Train Sample
  - Validation Sample
  - Testing Sample
- Input/Output
  - Features
  - Labels
- Fitting a model
  - Hypotheses
  - Hyperparameters
- Error Analysis
  - Metrics

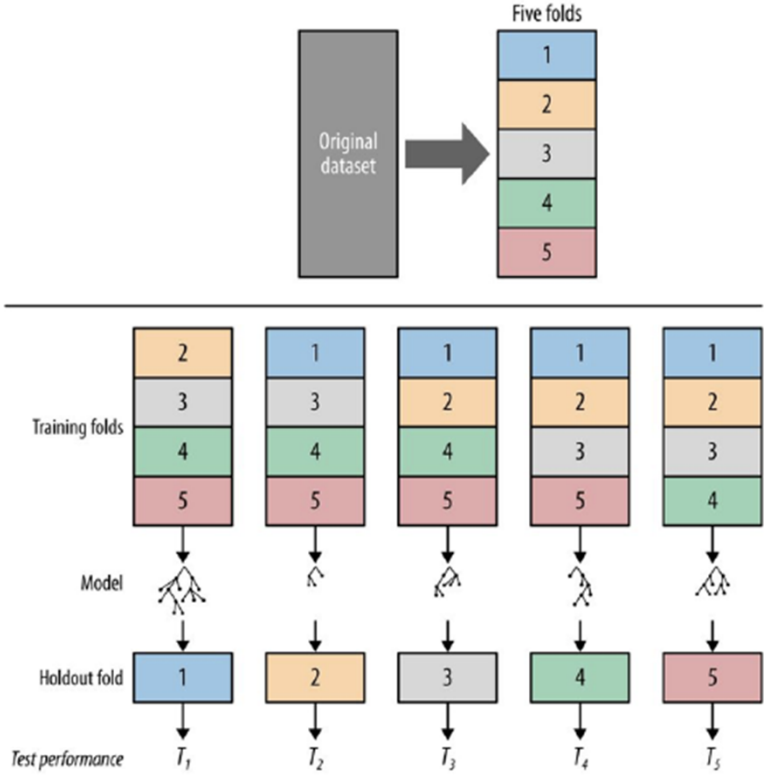
Components of Machine Learning Application



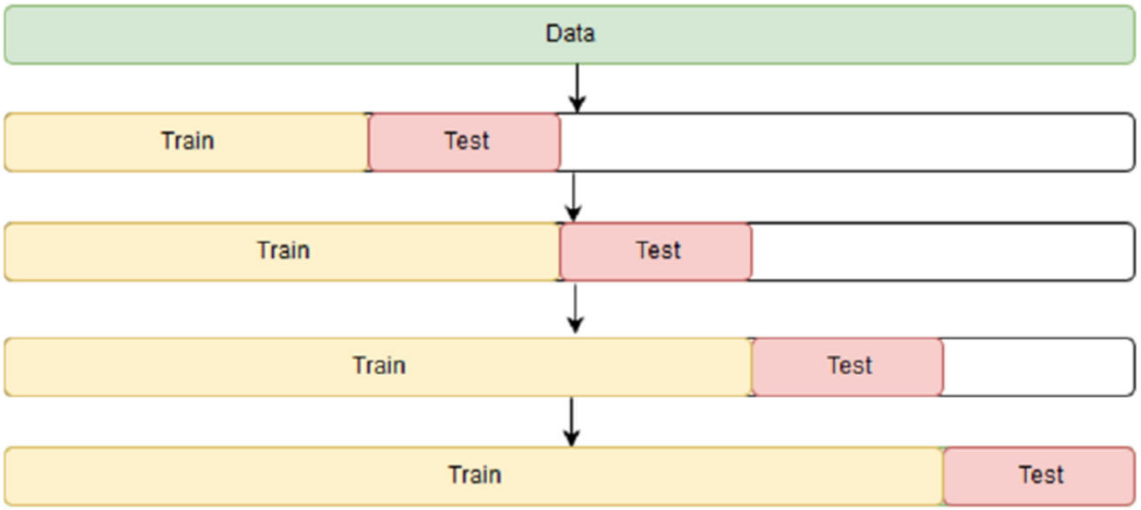
Components of Machine Learning Application



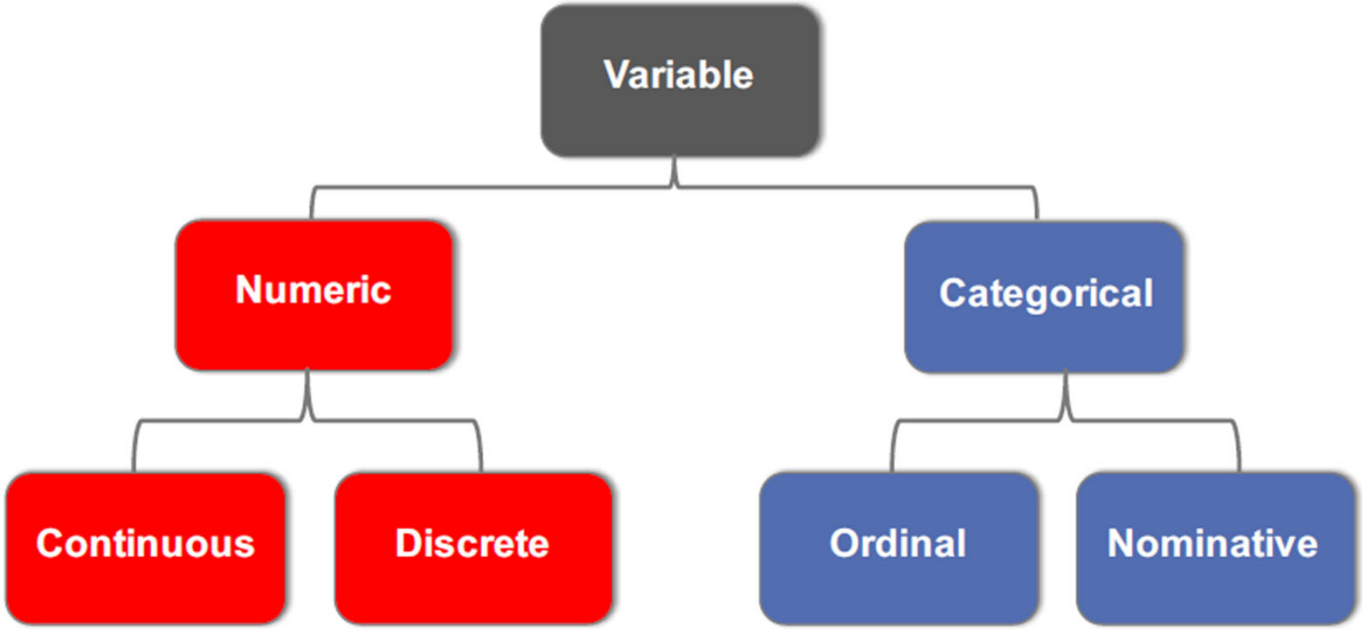
Train, Validate, Test



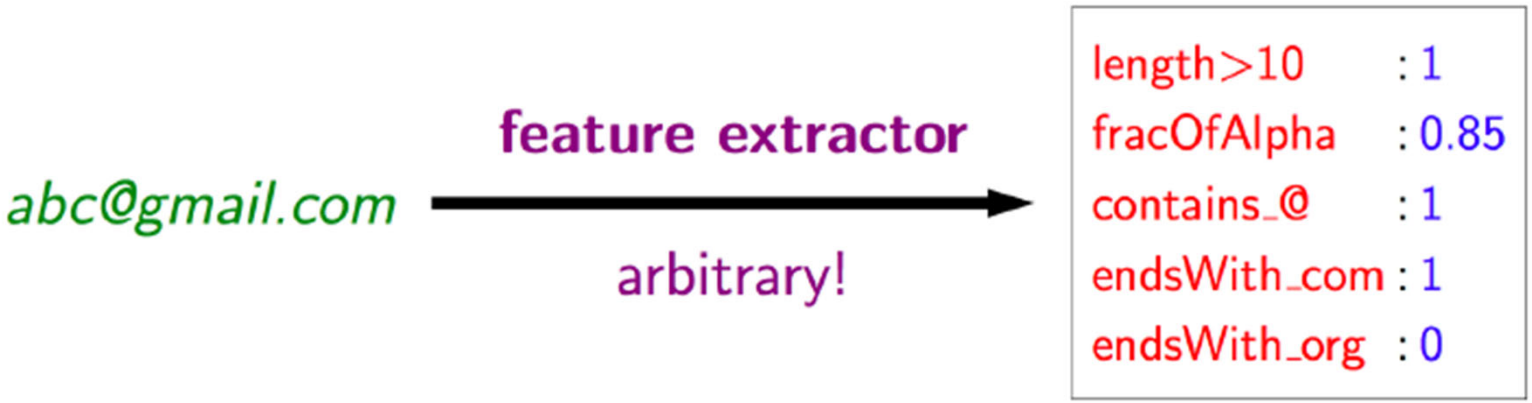
Train, Validate, Test



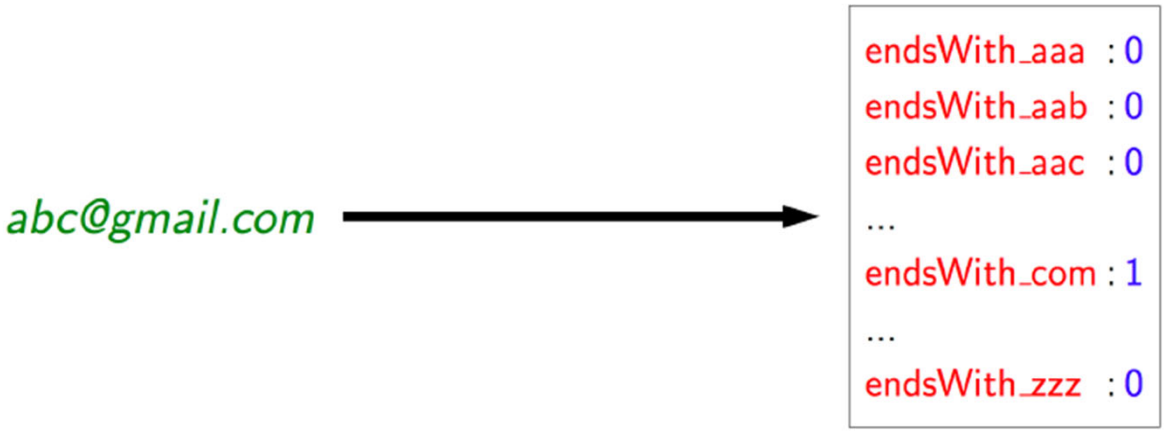
Features and Labels



Features and Labels



Features and Labels



Features and Labels

Array representation (good for dense features):

```
[0.85, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0]
```

Map representation (good for sparse features):

```
{"fracOfAlpha": 0.85, "contains_@": 1}
```

Features and Labels

> Anyone knows how much it costs to host a web portal ?
> Well, it depends on how many visitors youre expecting. This can be anywhere from less than 10 bucks a month to a couple of $100. You should checkout http://www.rackspace.com/ or perhaps Amazon EC2 if youre running something big..

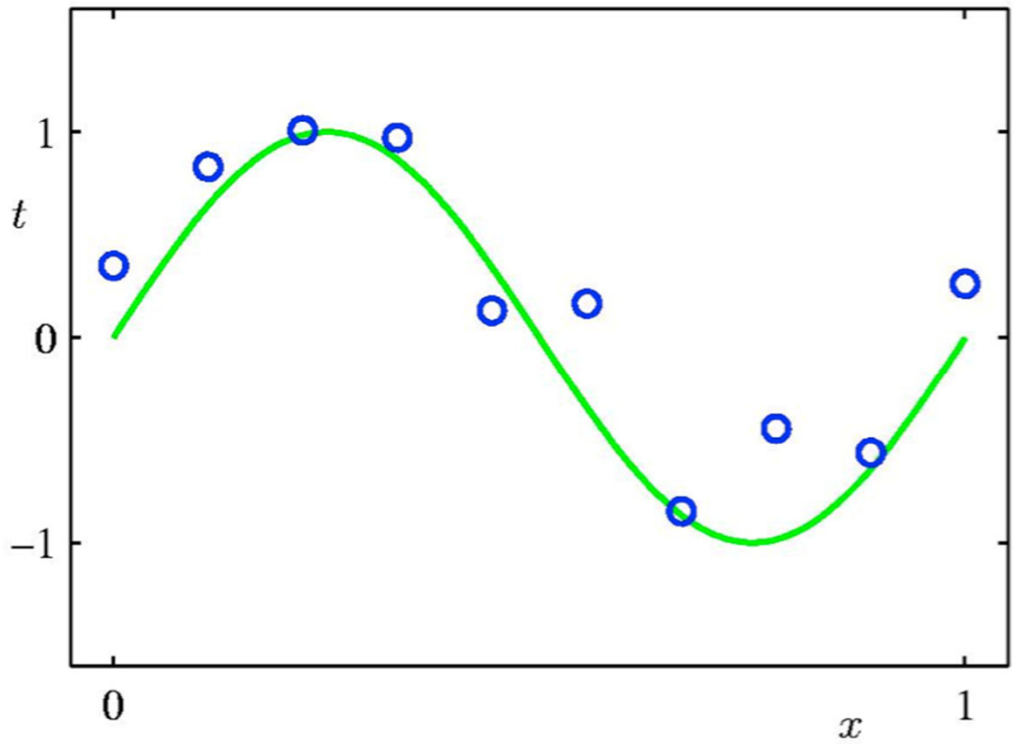
To unsubscribe yourself from this mailing list, send an email to: groupname-unsubscribe@egroups.com

Figure 1: Sample e-mail in SpamAssassin corpus before pre-processing.

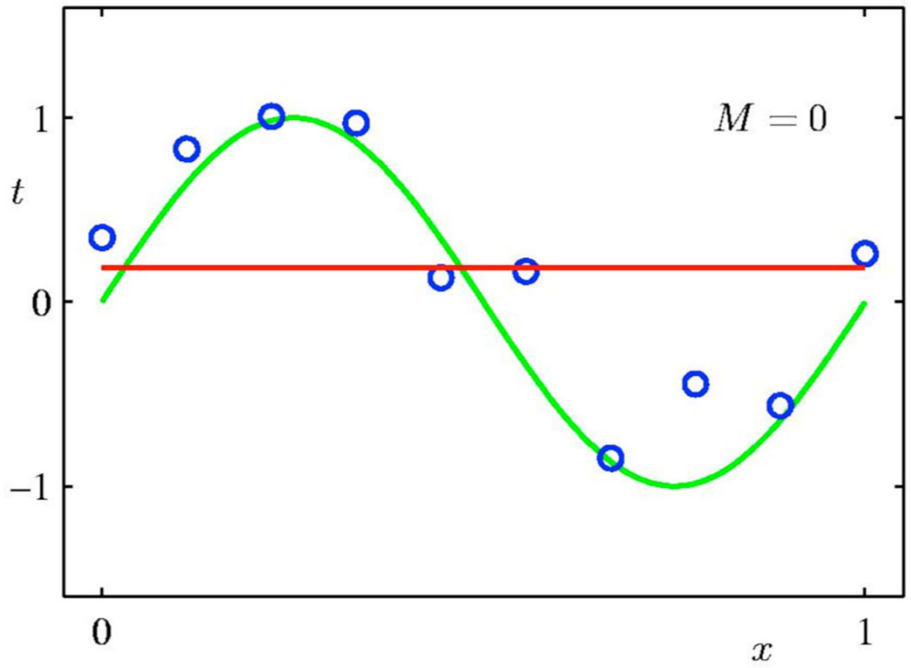
anyon know how much it cost to host a web portal well it depend on how mani visitor your expect thi can be anywher from less than number buck a month to a coupl of dollarnumb you should checkout httpaddr or perhap amazon ecnumb if your run someth big to unsubscrib yourself from thi mail list send an email to emailaddr

Figure 2: Pre-processed version of the sample e-mail from Figure 1.

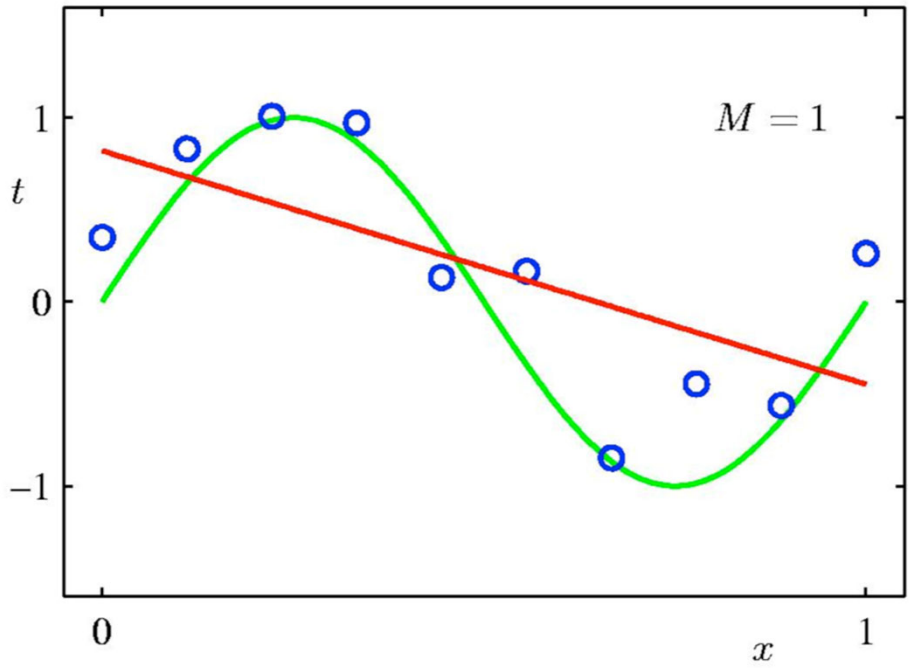
Hypotheses and Hyperparameters



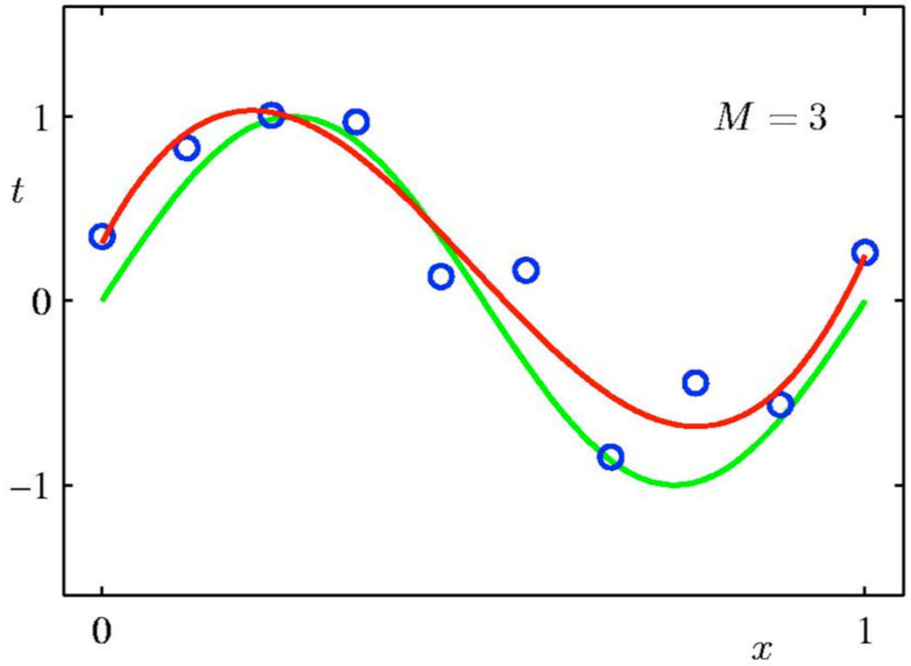
Hypotheses and Hyperparameters



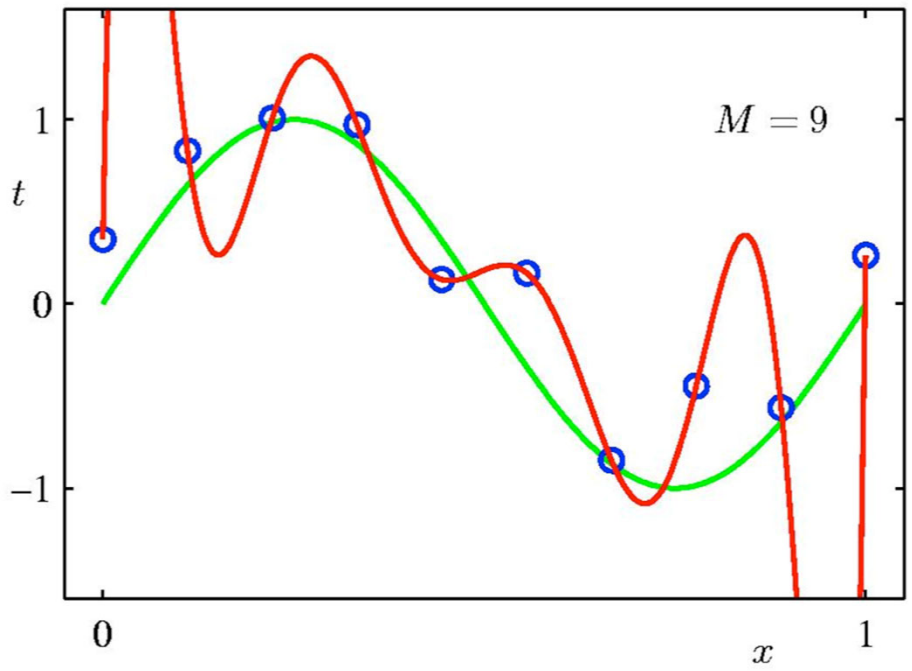
Hypotheses and Hyperparameters



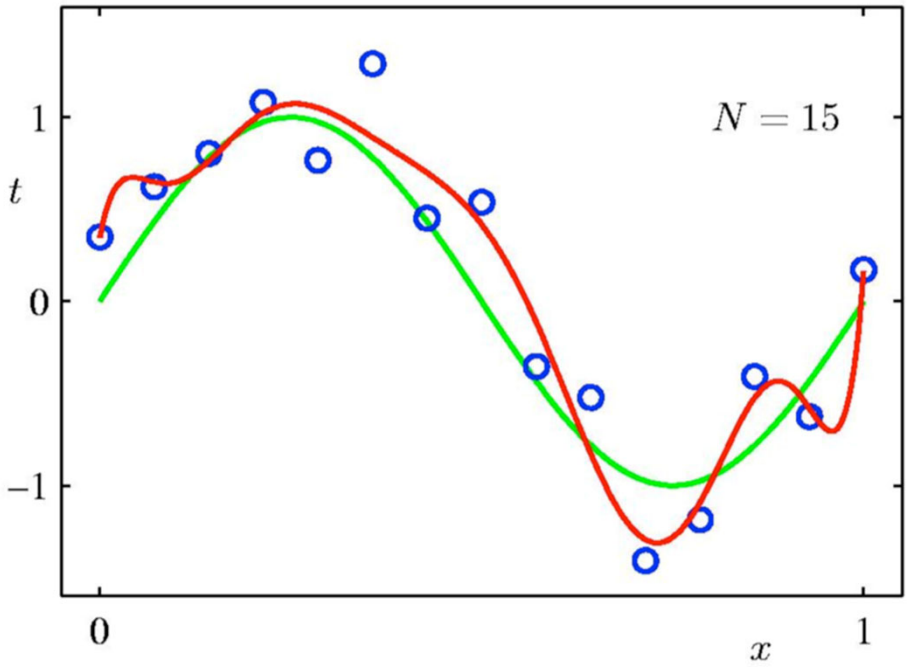
Hypotheses and Hyperparameters



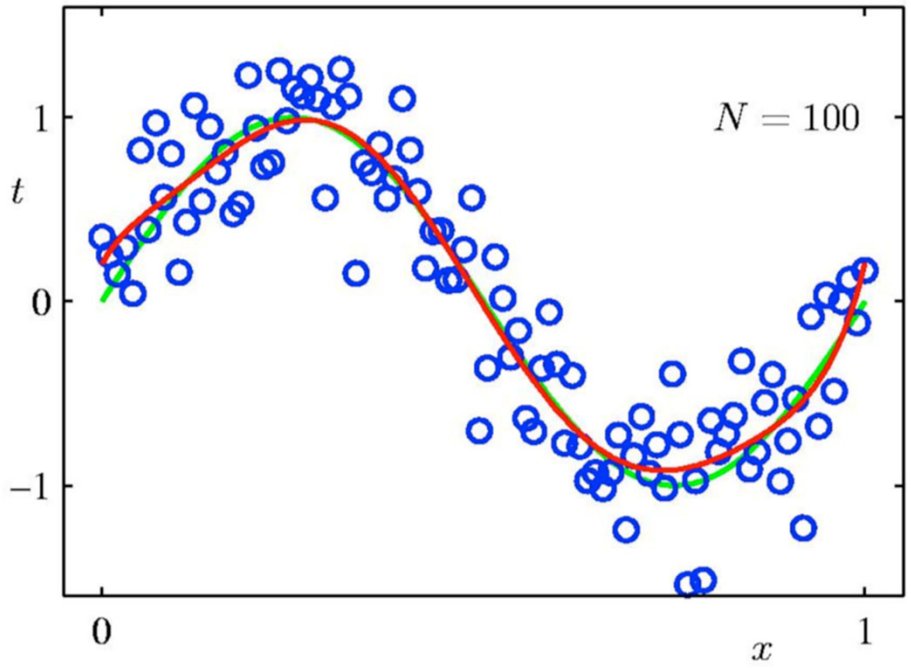
Hypotheses and Hyperparameters



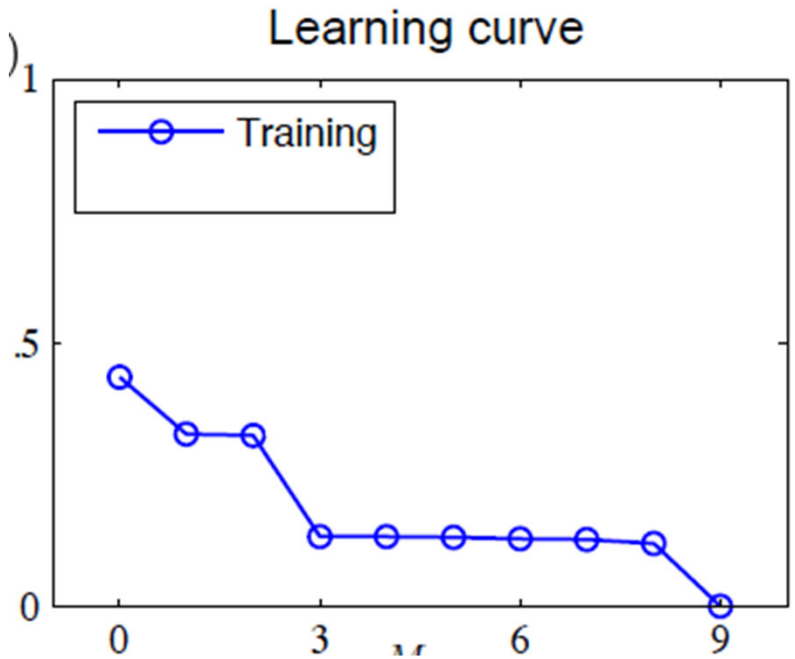
Hypotheses and Hyperparameters



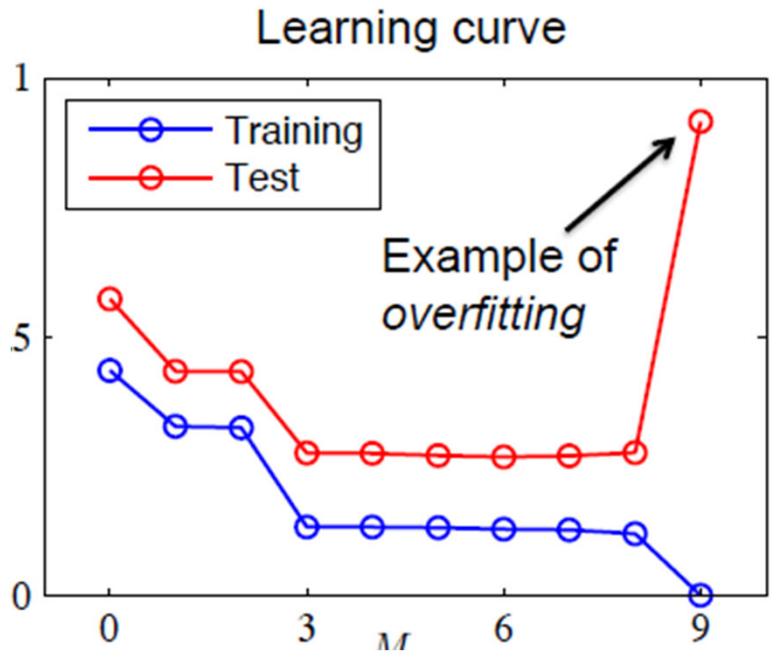
Hypotheses and Hyperparameters



Hypotheses and Hyperparameters



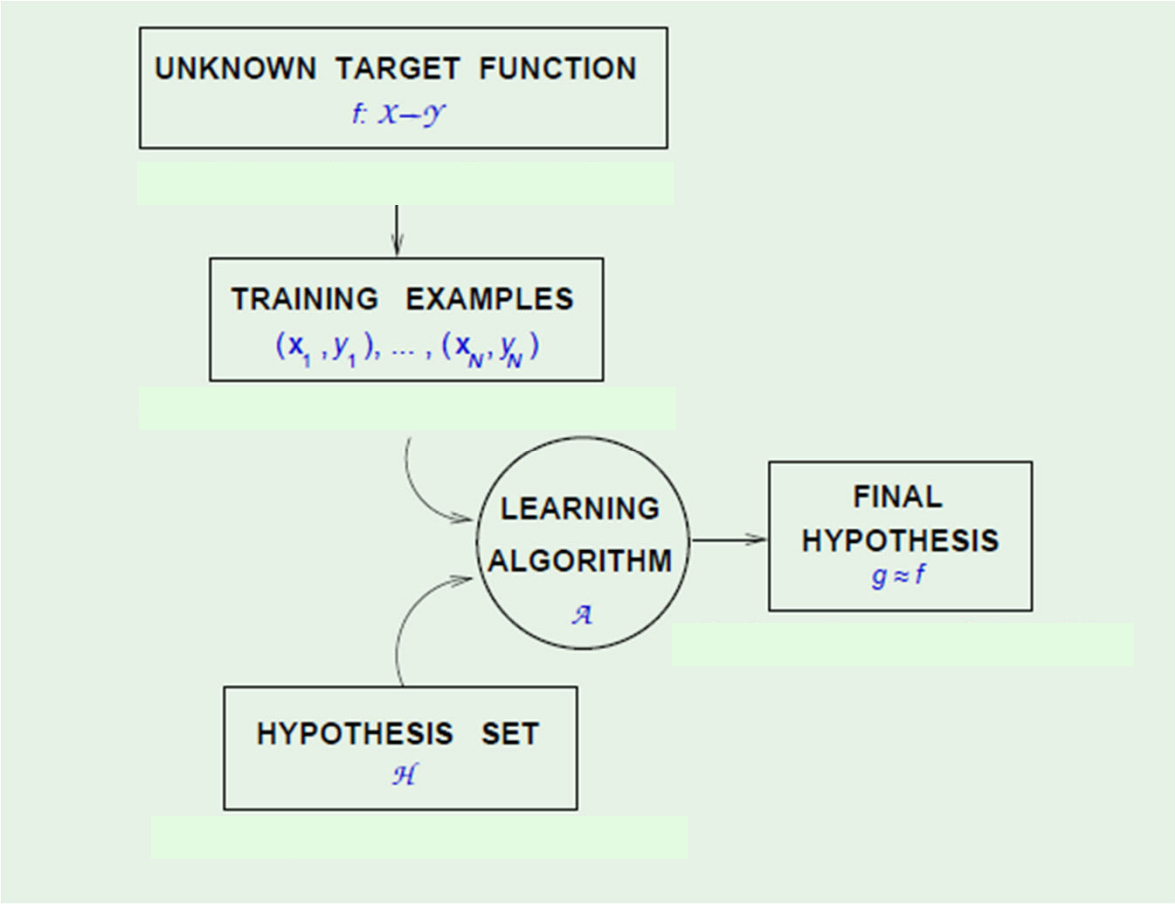
Hypotheses and Hyperparameters



Metrics

		True condition	
	Total population	Condition positive	Condition negative
Predicted condition	Predicted condition positive	**True positive**	**False positive**, Type I error
	Predicted condition negative	**False negative**, Type II error	**True negative**

Lesson


UNKNOWN TARGET FUNCTION
$f: X \to Y$
TRAINING EXAMPLES
$(x_1, y_1), \ldots, (x_N, y_N)$
LEARNING ALGORITHM
$\mathcal{A}$
FINAL HYPOTHESIS
$g \approx f$
HYPOTHESIS SET
$\mathcal{H}$

Lesson

- Step 1 (Input)

$$e \mapsto \mathbf{f}(e) = (h_1(e), \ldots, h_N(e)) = (x_1, \ldots, x_N)$$

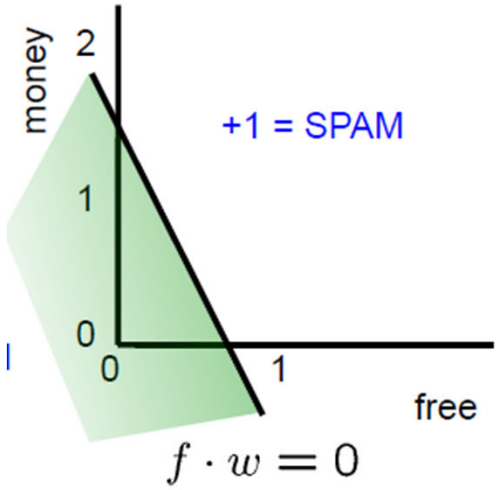
- Step 2 (Combine)

$$\langle \mathbf{w}, \mathbf{f}(e) \rangle = \sum_{i=1}^{N} w_i x_i$$

- Step 3 (Output)

$$\text{sign}(z) = \begin{cases} 1 & \text{if } z \geq 0 \\ -1 & \text{if } z < 0 \end{cases}$$

Lesson



x

"free money"

$f(x)$

BIAS	1
free	1
money	1
...	

w

BIAS	-3
free	4
money	2
...	

$$(1)(-3) + (1)(4) + (1)(2) + \ldots = 3$$

Lesson

- Hypothesis

$$\text{sign}\left(\langle \mathbf{w}, \mathbf{f}(e)\rangle - \text{threshold}\right) = \begin{cases} 1 & \text{then spam} \\ -1 & \text{then not spam} \end{cases}$$

- Step 2 (Combine)

$$f_{N+1}(e) \equiv 1$$

$$w_{N+1} = -\text{threshold}$$

- Step 3 (Output)

$$\text{sign}\left(\langle \mathbf{w}, \mathbf{f}(e)\rangle\right) = \begin{cases} 1 & \text{then spam} \\ -1 & \text{then not spam} \end{cases}$$

Lesson

input: A training set $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)$

initialize: $\mathbf{w}^{(1)} = (0, \ldots, 0)$

  for $t = 1, 2, \ldots$

    if $(\exists\, i \;\text{ s.t. }\; y_i \langle \mathbf{w}^{(t)}, \mathbf{x}_i \rangle \leq 0)$ then

      $\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} + y_i \mathbf{x}_i$

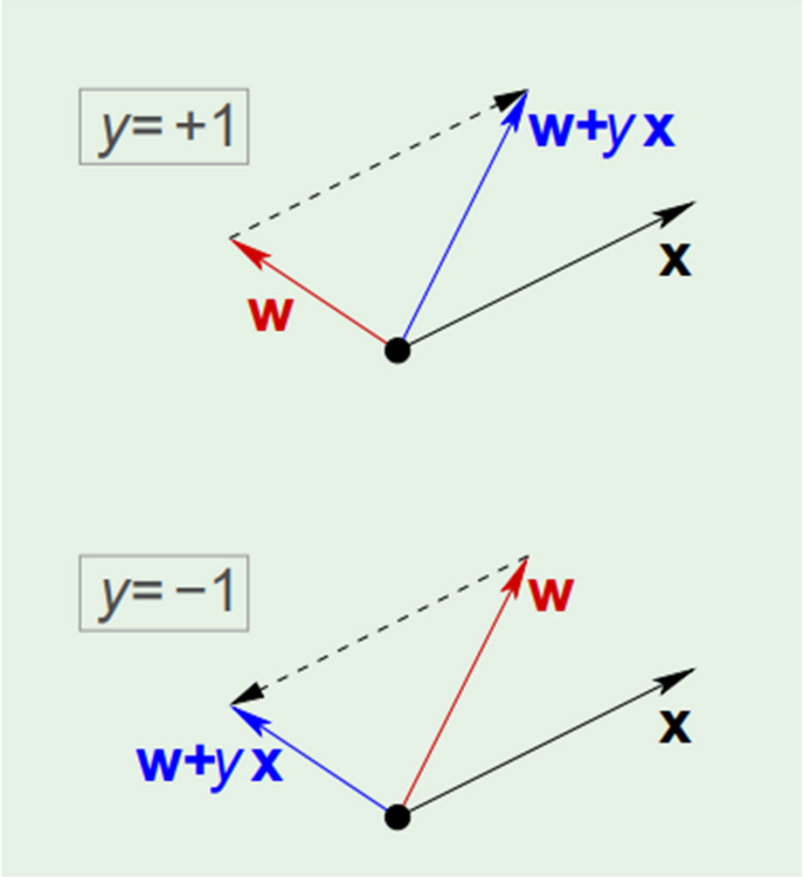
    else

      output $\mathbf{w}^{(t)}$

Lesson

$$y \langle w_t, x \rangle$$

$$y \langle w_{t+1}, x \rangle = y \langle w_t, x \rangle + ||x||^2$$



Lesson

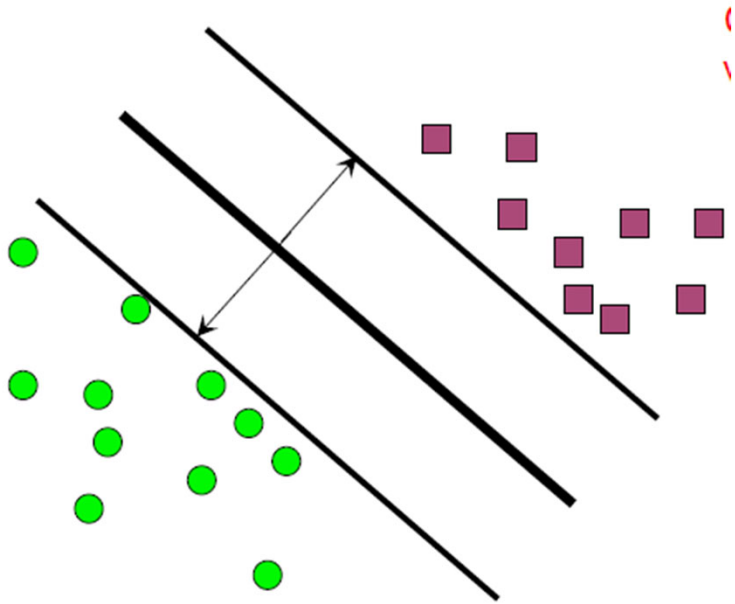
- Set

$$\gamma = \min_{1 \leq i \leq N} \frac{y_i \langle w, x_i \rangle}{||w||}$$

$$R = \max_{1 \leq i \leq N} ||x_i||$$

- The maximum number of mistakes made by the perceptron algorithm is $\frac{R^2}{\gamma^2}$

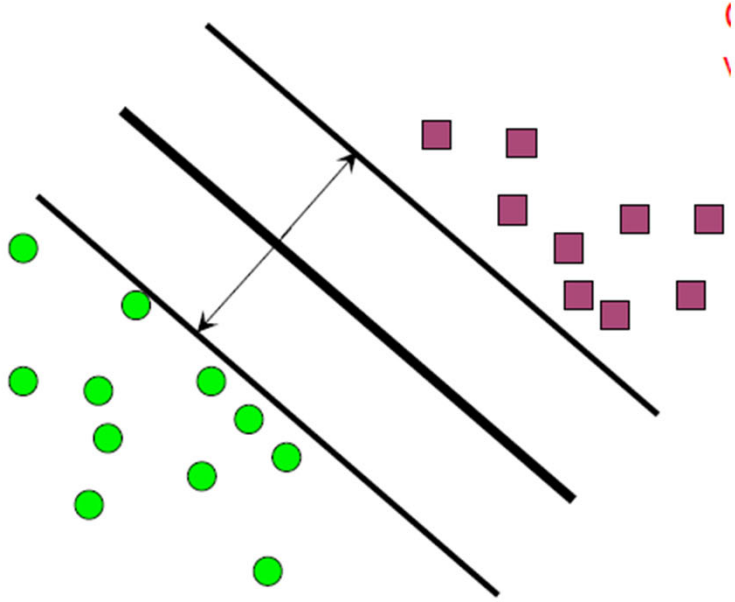
Lesson



Lesson

The distance between x and the plane defined by (w, b) is

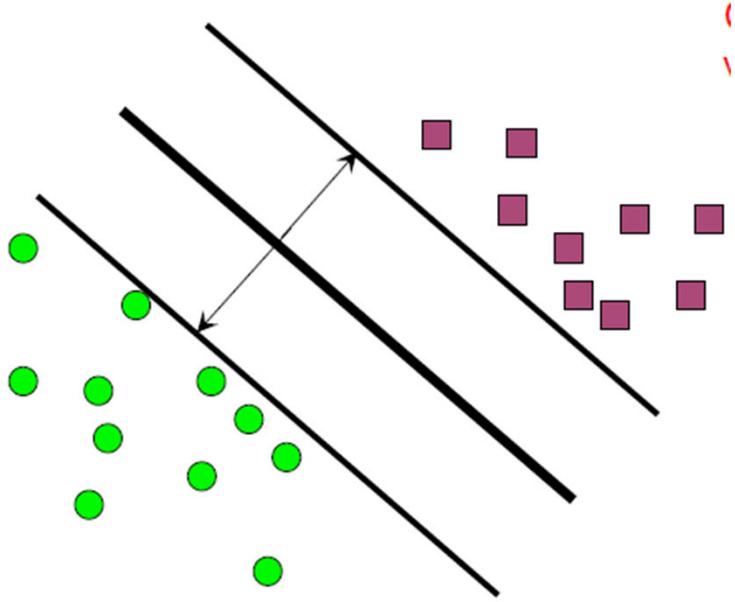
$$\frac{|\langle w, x\rangle + b|}{||w||}$$



Lesson

The distance between x and the plane defined by (w, b) is

$$\frac{|\langle w, x\rangle + b|}{||w||}$$



Lesson

Determine

$$\min\{\|\mathbf{x}-\mathbf{v}\| : \langle \mathbf{w}, \mathbf{v}\rangle + b = 0\}.$$

Set

$$\mathbf{v} = \mathbf{x} - (\langle \mathbf{w}, \mathbf{x}\rangle + b)\mathbf{w}$$

Note

$$\langle \mathbf{w}, \mathbf{v}\rangle + b = \langle \mathbf{w}, \mathbf{x}\rangle - (\langle \mathbf{w}, \mathbf{x}\rangle + b)\|\mathbf{w}\|^2 + b = 0,$$

$$\|\mathbf{x}-\mathbf{v}\| = |\langle \mathbf{w}, \mathbf{x}\rangle + b|\,\|\mathbf{w}\| = |\langle \mathbf{w}, \mathbf{x}\rangle + b|.$$

Lesson

For an u such that

$$\langle \mathbf{w}, \mathbf{u} \rangle + b = 0.$$

have

$$\begin{aligned}
\|\mathbf{x} - \mathbf{u}\|^2 &= \|\mathbf{x} - \mathbf{v} + \mathbf{v} - \mathbf{u}\|^2 \\
&= \|\mathbf{x} - \mathbf{v}\|^2 + \|\mathbf{v} - \mathbf{u}\|^2 + 2\langle \mathbf{x} - \mathbf{v}, \mathbf{v} - \mathbf{u} \rangle \\
&\geq \|\mathbf{x} - \mathbf{v}\|^2 + 2\langle \mathbf{x} - \mathbf{v}, \mathbf{v} - \mathbf{u} \rangle \\
&= \|\mathbf{x} - \mathbf{v}\|^2 + 2(\langle \mathbf{w}, \mathbf{x} \rangle + b)\langle \mathbf{w}, \mathbf{v} - \mathbf{u} \rangle \\
&= \|\mathbf{x} - \mathbf{v}\|^2,
\end{aligned}$$

Lesson

- Advantages
  - Bound on Number of Iterations Independent of Dimension
  - On-Line Learning
- Disadvantages
  - Limited to Linearly Separable
  - Overfitting (Improved by Averaging or Early Stopping)

Demo

The Data

1. The historical data consist of 39,859 customers. The historical data contain 19,901 customers that churned (*i.e.* left the company) and 19,958 that did not churn (see the "churndep" variable).
3. Here are the data set's 11 possible predictor variables for churning behavior:

Position	Variable Name	Variable Description
1	revenue	Mean monthly revenue in dollars
2	outcalls	Mean number of outbound voice calls
3	incalls	Mean number of inbound voice calls
4	months	Months in Service
5	eqpdays	Number of days the customer has had his/her current equipment
6	webcap	Handset is web capable
7	marryyes	Married (1=Yes; 0=No)
8	travel	Has traveled to non-US country (1=Yes; 0=No)
9	pcown	Owns a personal computer (1=Yes; 0=No)
10	credited	Possesses a credit card (1=Yes; 0=No)
11	retcalls	Number of calls previously made to retention team

The dependent variable, Churndep, = 1 if the customer churned, = 0 otherwise.