

Aprendizagem de Máquina

Francisco Carvalho e Leandro Almeida

UFPE - Universidade Federal de Pernambuco

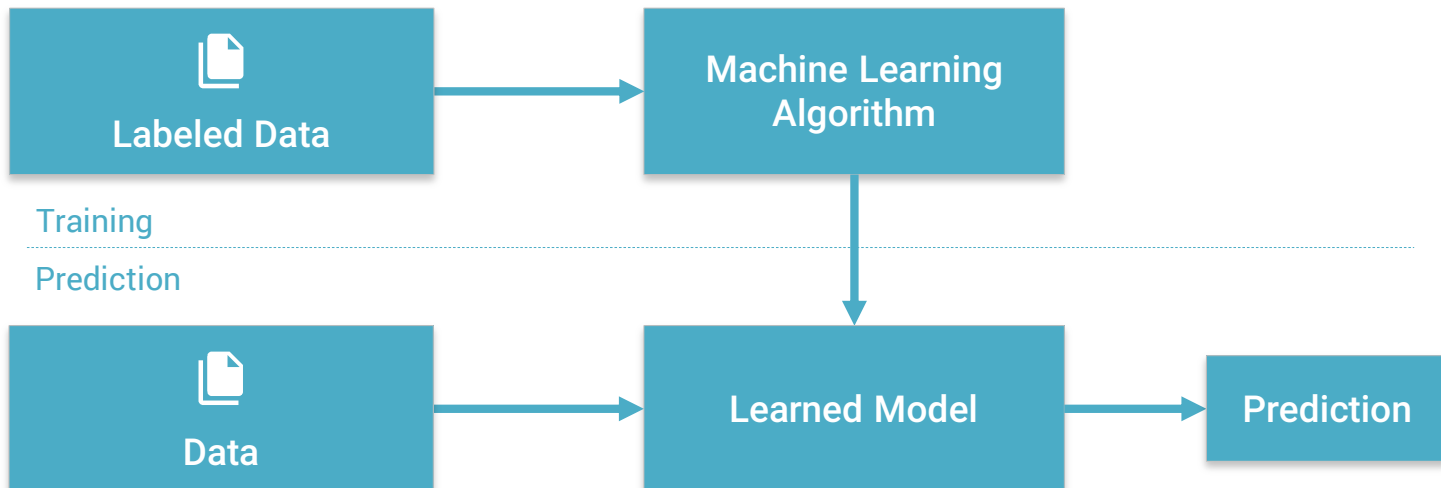
CIn - Centro de Informática

Machine Learning - Basics

Introduction



Machine Learning is a type of Artificial Intelligence that provides computers with the ability to **learn without being explicitly programmed**.



Provides **various techniques** that can learn from and make predictions on data

Machine Learning - Basics

Learning Approaches



Supervised Learning: Learning with a **labeled training set**
Example: email spam detector with training set of already labeled emails



Unsupervised Learning: **Discovering patterns** in unlabeled data
Example: cluster similar documents based on the text content



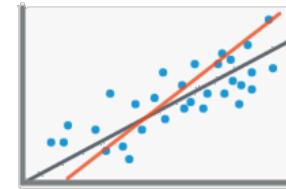
Reinforcement Learning: learning based on **feedback** or reward
Example: learn to play chess by winning or losing

Machine Learning - Basics

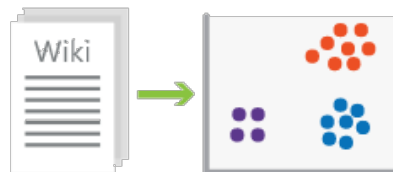
Problem Types



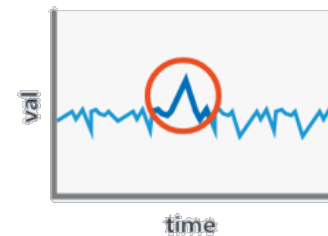
Classification
(supervised – predictive)



Regression
(supervised – predictive)



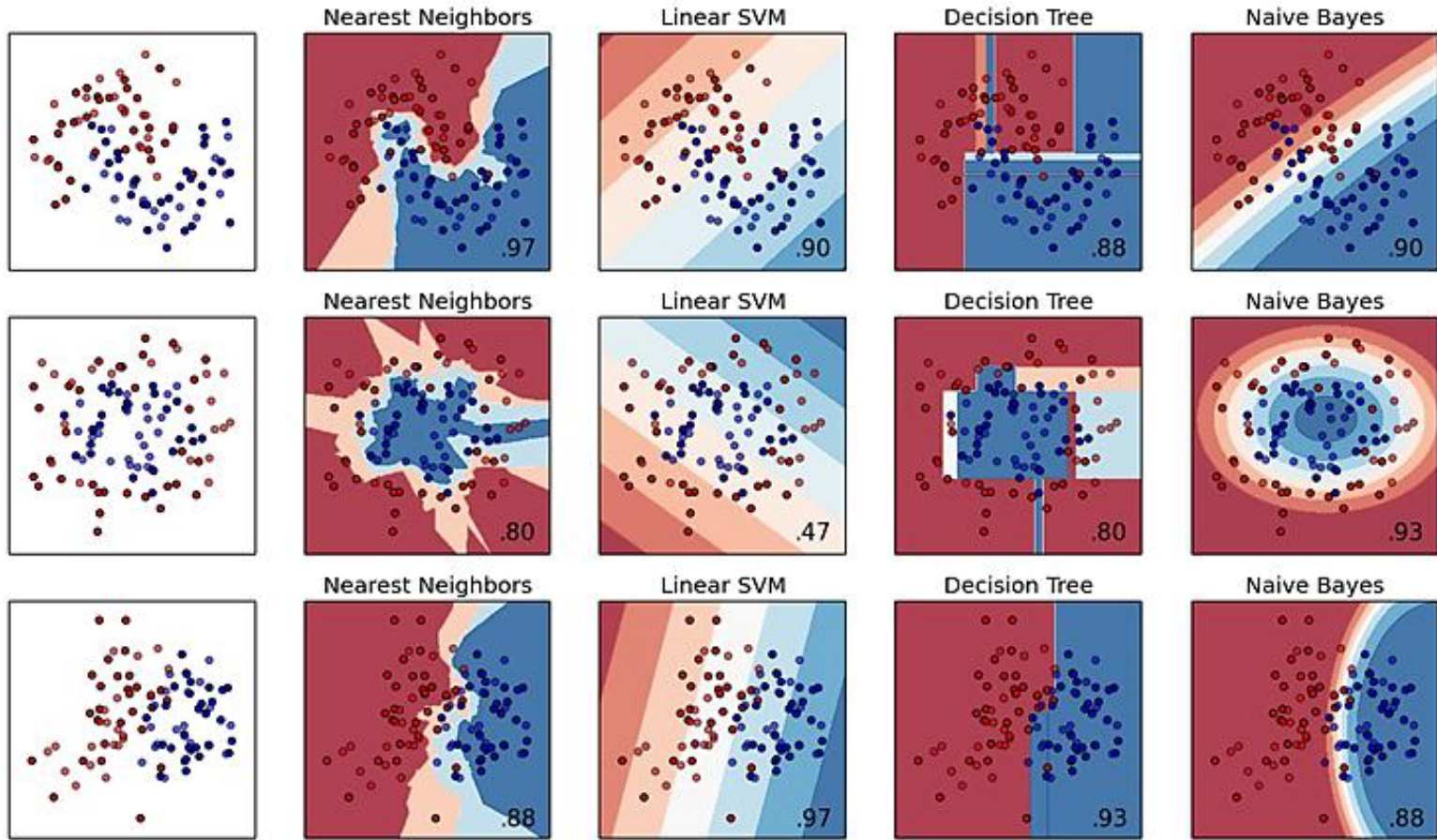
Clustering
(unsupervised – descriptive)



Anomaly Detection
(unsupervised – descriptive)

Machine Learning - Basics

Algorithms Comparison - Classification





The Home of Data Science

COMPETITIONS • DATASETS • CUSTOMER SOLUTIONS • JOBS BOARD

Active Competitions

Active Competitions

All Competitions



Second Annual Data Science Bowl

Transforming How We Diagnose Heart Disease

7.3 days
755 teams
\$200,000



Santander Customer Satisfaction

Which customers are happy customers?

56 days
639 teams
226 scripts
\$60,000



Home Depot Product Search Relevance

Predict the relevance of search results on homedepot.com

49 days
1379 teams
1007 scripts
\$40,000



BNP Paribas Cardif Claims Management

Can you accelerate BNP Paribas Cardif's claims management process?

42 days
1820 teams
690 scripts
\$30,000



March Machine Learning Mania 2016

Predict the 2016 NCAA Basketball Tournament

5.3 days
522 teams
327 scripts
\$25,000



Yelp Restaurant Photo Classification

36 days
169 teams

Componentes do aprendizado



Metafora: Aprovação de crédito

Applicant information:

age	23 years
gender	male
annual salary	\$30,000
years in residence	1 year
years in job	1 year
current debt	\$15,000
...	...

Aprovar crédito?

Learning From Data, Prof. Yaser Abu-Mostafa



UNIVERSIDADE FEDERAL
DE PERNAMBUCO

cin.ufpe.br

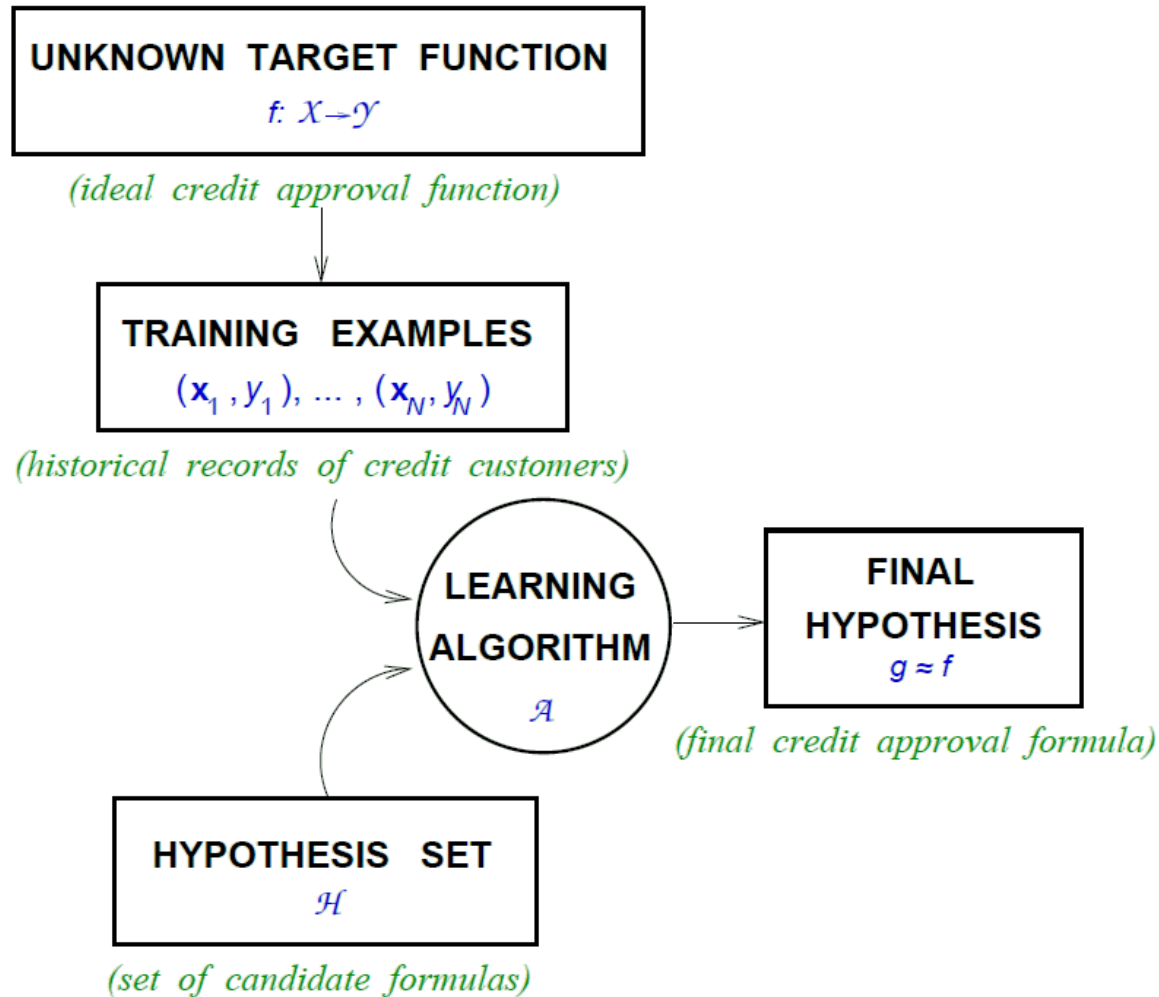
Formalization:

- Input: $\mathbf{x}$ (*customer application*)
- Output: y (*good/bad customer?*)
- Target function: $f : \mathcal{X} \rightarrow \mathcal{Y}$ (*ideal credit approval formula*)
- Data: $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)$ (*historical records*)



- Hypothesis: $g : \mathcal{X} \rightarrow \mathcal{Y}$ (*formula to be used*)

Componentes do aprendizado



Learning From Data, Prof. Yaser Abu-Mostafa

Componentes da solução

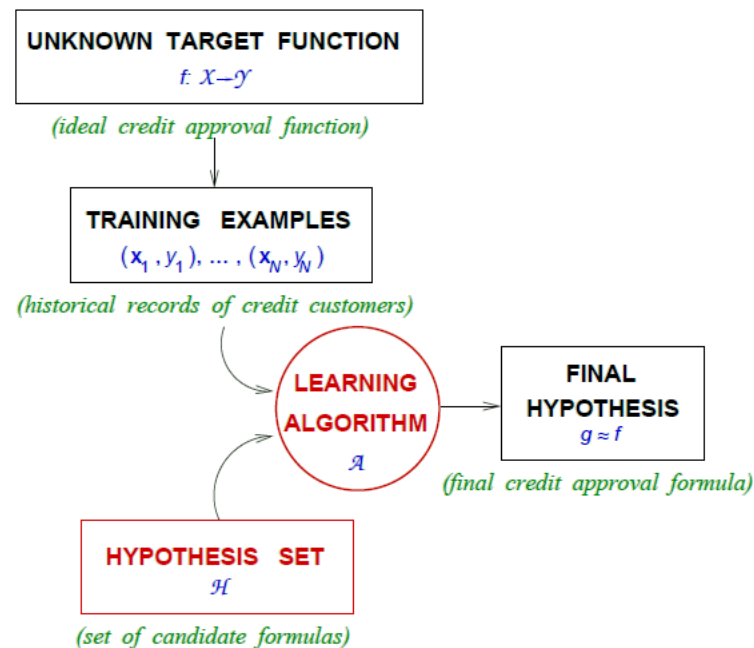
The 2 solution components of the learning problem:

- The Hypothesis Set

$$\mathcal{H} = \{h\} \quad g \in \mathcal{H}$$

- The Learning Algorithm

Together, they are referred to as the *learning model*.



Regiões de Separação - Linear



- Sabe-se que se formarmos uma **combinação linear de duas variáveis**, e igualá-la a um número, então os pontos no espaço bidimensional podem ser divididos em três categorias:

- a) pontos **pertencentes à linha** com coordenadas tais que

$$w_1 \cdot x_1 + w_2 \cdot x_2 = \theta$$

- b) pontos **em um lado da linha** tem coordenadas tais que

$$w_1 \cdot x_1 + w_2 \cdot x_2 < \theta$$

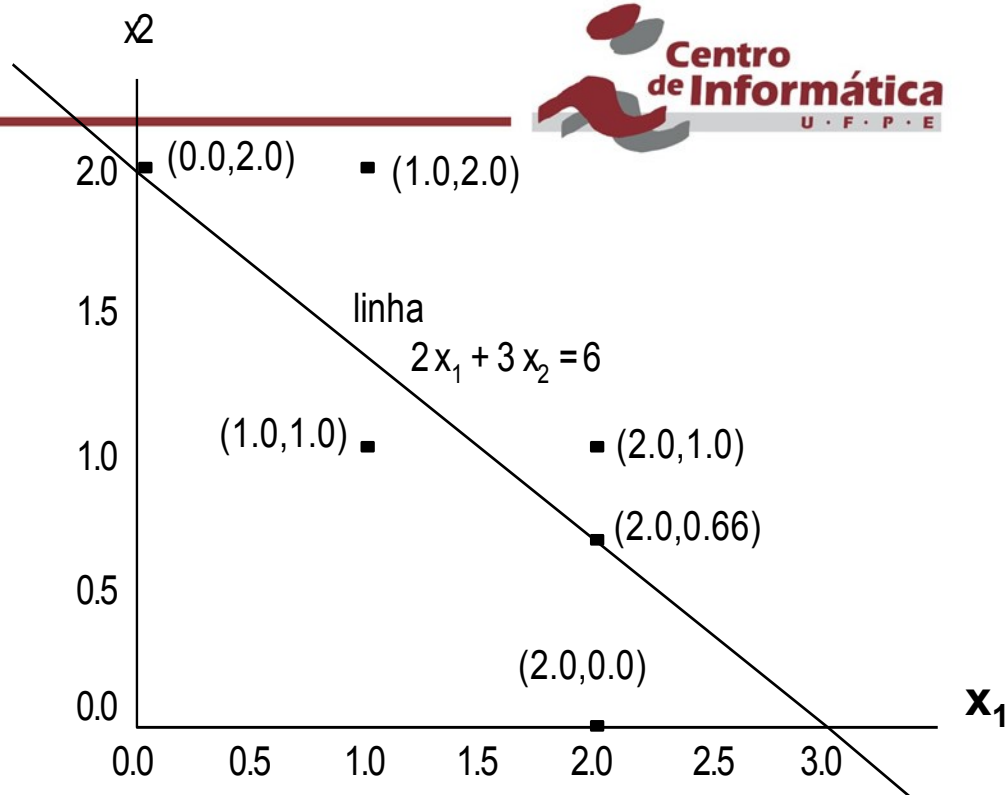
- c) pontos **no outro lado da linha** tem coordenadas tais que

$$w_1 \cdot x_1 + w_2 \cdot x_2 > \theta.$$



Exemplo

Pontos (x_1 , x_2)	$2x_1 + 3x_2$	posição
(0.0, 2.0)	6	linha
(1.0, 1.0)	5	abaixo
(1.0, 2.0)	8	acima
(2.0, 0.0)	4	abaixo
(2.0, 0.66)	6	linha
(2.0, 1.0)	7	acima



Posição dos pontos em função da linha $2x_1 + 3x_2 = 6$ de delimitação.

Linha: $2x_1 + 3x_2 = 6$

Acima: $2x_1 + 3x_2 > 6$

$\theta = 6$

Abaixo: $2x_1 + 3x_2 < 6$

A simple hypothesis set - the 'perceptron'

For input $\mathbf{x} = (x_1, \dots, x_d)$ 'attributes of a customer'

Approve credit if $\sum_{i=1}^d w_i x_i > \text{threshold},$

Deny credit if $\sum_{i=1}^d w_i x_i < \text{threshold}.$

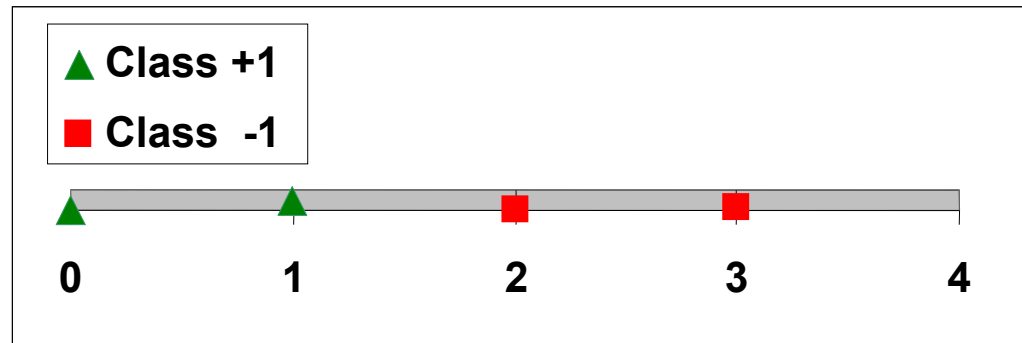
This linear formula $h \in \mathcal{H}$ can be written as

$$h(\mathbf{x}) = \text{sign} \left(\left(\sum_{i=1}^d w_i x_i \right) - \text{threshold} \right)$$

Exemplo

Vantagens da separabilidade linear

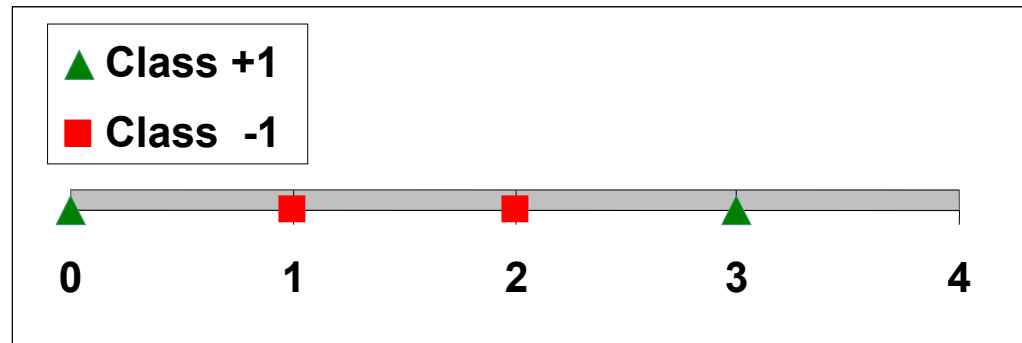
X_1	Class
0	+1
1	+1
2	-1
3	-1



Exemplo

Como separar as duas classes com apenas um ponto?

X_1	Class
0	+1
1	-1
2	-1
3	+1

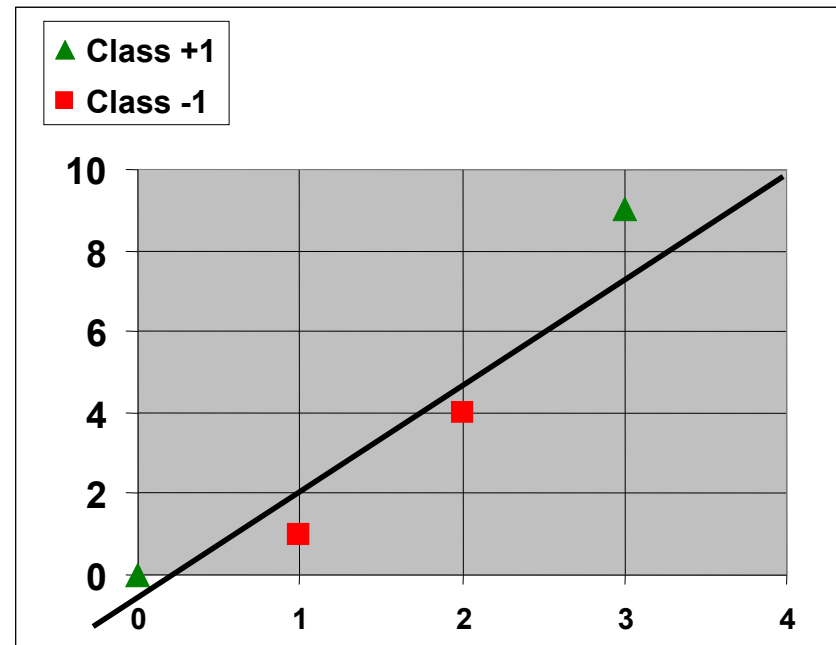


Exemplo

Possível solução é aumentar a quantidade de dimensões
(Teorema de Cover, 1965)

$$\Phi(X_1) = (X_1, X_1^2)$$

X_1	X_1^2	Class
0	0	+1
1	1	-1
2	4	-1
3	9	+1



A simple hypothesis set - the 'perceptron'

For input $\mathbf{x} = (x_1, \dots, x_d)$ 'attributes of a customer'

Approve credit if $\sum_{i=1}^d w_i x_i > \text{threshold},$

Deny credit if $\sum_{i=1}^d w_i x_i < \text{threshold}.$

This linear formula $h \in \mathcal{H}$ can be written as

$$h(\mathbf{x}) = \text{sign} \left(\left(\sum_{i=1}^d w_i x_i \right) - \text{threshold} \right)$$

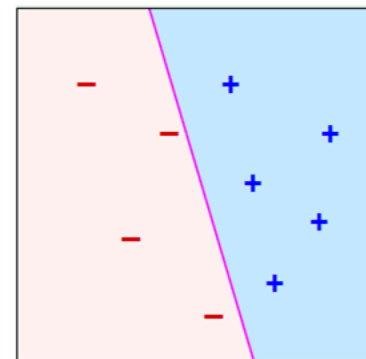
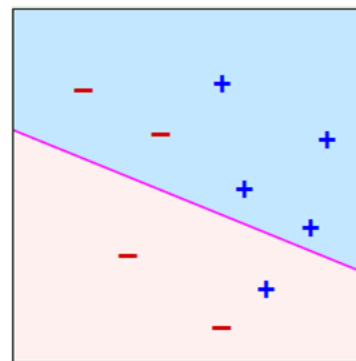
$$h(\mathbf{x}) = \text{sign} \left(\left(\sum_{i=1}^d w_i x_i \right) + w_0 \right)$$

Introduce an artificial coordinate $x_0 = 1$:

$$h(\mathbf{x}) = \text{sign} \left(\sum_{i=0}^d w_i x_i \right)$$

In vector form, the perceptron implements

$$h(\mathbf{x}) = \text{sign}(\mathbf{w}^T \mathbf{x})$$



'linearly separable' data

A simple learning algorithm - PLA

The perceptron implements

$$h(\mathbf{x}) = \text{sign}(\mathbf{w}^T \mathbf{x})$$

Given the training set:

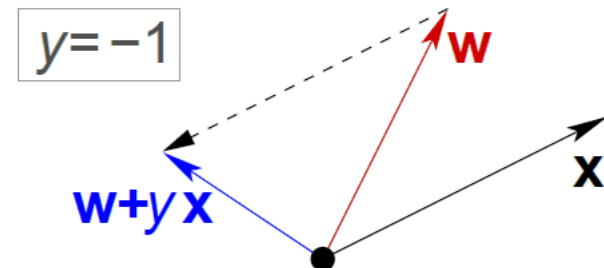
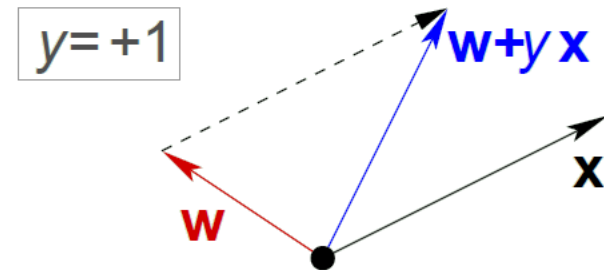
$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)$$

pick a **misclassified** point:

$$\text{sign}(\mathbf{w}^T \mathbf{x}_n) \neq y_n$$

and update the weight vector:

$$\mathbf{w} \leftarrow \mathbf{w} + y_n \mathbf{x}_n$$



Iterations of PLA

- One iteration of the PLA:

$$\mathbf{w} \leftarrow \mathbf{w} + y\mathbf{x}$$

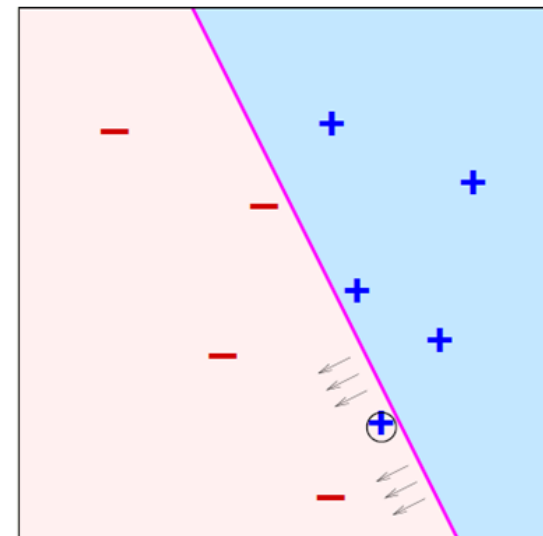
where $(\mathbf{x}, y)$ is a misclassified training point.

- At iteration $t = 1, 2, 3, \dots$, pick a misclassified point from

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)$$

and run a PLA iteration on it.

- That's it!



* Bias data?



Possibilidades



GTC 2019

<https://www.youtube.com/watch?v=GWL1HNHDSq4>

GTC 2018

<https://www.youtube.com/watch?v=GiZ7kyrwZGQ>

GTC 2017

<https://www.youtube.com/watch?v=SUNPrR4o5ZA>

<https://www.youtube.com/watch?v=Dy0hJWltsyE&list=RDSUNPrR4o5ZA&index=2>



- Em geral, é difícil articular o conhecimento que precisamos para construir um sistema de Inteligência Computacional
- Na verdade, algumas vezes, não temos nem este conhecimento
- Em alguns casos, podemos construir sistemas em que eles mesmos aprendam o conhecimento necessário

■ Principal preocupação

- Construção de programas de computador que melhoram seu desempenho por meio de experiência

■ Técnicas orientadas a dados

- Aprendem automaticamente a partir de grandes volumes de dados
- Geração de hipóteses a partir dos dados

Aprendizado de Máquina (Exemplos)



- AM tem sido utilizado rotineiramente para descobrir informações valiosas de grandes bases contendo:
 - Registros de manutenção de equipamentos
 - Aplicações para empréstimos
 - Transações financeiras
 - Registros médicos
 - Genomas



Inferência Indutiva (1/2)



▪ Indução

- Um processo de raciocínio para uma conclusão sobre todos os membros de uma classe por meio do exame de apenas uns poucos membros da classe
- De maneira geral, raciocínio do particular para o geral
- Por exemplo, se eu noto que:
 - Todos os pacientes com Déficit de Atenção atendidos em 1986 sofriam de Ansiedade
 - Todos os pacientes com Déficit de Atenção atendidos em 1987 sofriam de Ansiedade
 - ...
 - **Posso inferir logicamente que Todos os pacientes que sofrem de Déficit de Atenção também sofrem de Ansiedade**
 - Isto pode ser ou não verdade, mas propicia uma boa **generalização**



Inferência Indutiva (2/2)



- De uma maneira mais “formal”...
 - Para um conjunto de objetos, $X=\{a,b,c,d,\dots\}$, se a propriedade P é verdade para a , e se P é verdade para b , e se P é verdade para c ,... então P é verdade para todo X
 - O conhecimento novo baseado em vários casos (indução) é geralmente verdadeiro desde que os sistemas estudados sejam bem comportados
 - Se o número de objetos (exemplos) for insuficiente, ou se não forem bem escolhidos, as hipóteses obtidas podem ser de pouco valor
 - A inferência indutiva é um dos principais métodos utilizados para derivar conhecimento novo e prever eventos futuros

Um programa aprende a partir da experiência **E**, em relação a uma classe de tarefas **T**, com medida de desempenho **P**, se seu desempenho em **T**, medido por **P**, melhora com **E**

Mitchell, 1997

Também chamado de **Aprendizado Indutivo**

- Detecção de bons clientes para um cartão de crédito
 - **Tarefa T**: classificar potenciais novos clientes como bons ou maus pagadores
 - **Medida de Desempenho P**: porcentagem de clientes classificados corretamente
 - **Experiência de Treinamento E**: uma base de dados histórica em que os clientes já conhecidos são previamente classificados como bons ou maus pagadores

- Reconhecimento de caracteres manuscritos
 - **Tarefa T**: reconhecer e classificar caracteres manuscritos
 - **Medida de Desempenho P**: percentagem de caracteres classificados corretamente
 - **Experiência de Treinamento E**: base de dados de caracteres manuscritos com a respectiva classificação

■ Técnicas de AM

- Ajustam seus parâmetros em uma **fase de treinamento**
 - Utiliza **conjunto** de dados de **treinamento**
- Avaliam seu desempenho para novos dados em, uma **fase de teste** (ou validação)
 - Utiliza **conjunto** de dados de **teste**
 - Desempenho depende da representatividade dos exemplos
 - Aprendizado é mais confiável quando exemplos de treinamento seguem uma **distribuição** semelhante à dos **exemplos de teste**

Paradigmas de AM



- Simbólico
- Baseado em Instâncias
- Estatístico
- Conexionista
- Evolucionário



- Constrói representações simbólicas de um conceito através da análise de seus exemplos e contra-exemplos

- Representações simbólicas geralmente assumem a forma de:
 - Expressão lógica
 - Árvore de decisão
 - Regras de produção
 - Rede semântica

- Classificam novos dados utilizando dados similares cuja classe é conhecida
- Análogo ao que fazemos para resolver novos problemas
- Exemplo: Raciocínio Baseado em Casos e k-vizinhos mais próximos

- Constrói um modelo estatístico do problema, geralmente utilizando a **regra de Bayes**
- Podem ser:
 - Paramétricas
 - Não paramétricas
 - Semi-paramétricas

- Computação Evolucionária (ou evolutiva)
 - Sistemas para a resolução de problemas que utilizam modelos computacionais baseados na teoria da evolução natural das espécies
 - Também chamados de algoritmos evolucionários
 - Inclui os **Algoritmos Genéticos**

- Redes Neurais Artificiais
 - Estruturas distribuídas
 - Formadas por um grande número de unidades de processamento conectadas entre si

■ Aprendizado Supervisionado

- O algoritmo de aprendizado (indutor) recebe um conjunto de exemplos de treinamento para os quais os rótulos da classe associada são conhecidos
- Cada exemplo (instância ou padrão) é descrito por um vetor de valores (atributos) e pelo rótulo da classe associada
- O objetivo do indutor é construir um classificador que possa determinar corretamente a classe de novos exemplos ainda não rotulados
- Para rótulos de classe discretos, esse problema é chamado de classificação e para valores contínuos como regressão

Tipos de Aprendizado de Máquina (2/3)



- Aprendizado Não-Supervisionado
 - O indutor analisa os exemplos fornecidos e tenta determinar se alguns deles podem ser agrupados de alguma maneira, formando agrupamentos ou clusters
 - Após a determinação dos agrupamentos, em geral, é necessário uma análise para determinar o que cada agrupamento significa no contexto problema sendo analisado



Conceitos e Definições de Aprendizado

Conceitos e Definições



- **Exemplo** (caso, registro ou dado) é uma tupla de valores de atributos
 - Um paciente, dados médicos sobre uma determinada doença
- **Atributo**: descreve uma característica ou um aspecto de um exemplo
 - Nominal: cor
 - Contínuo: peso
- **Classe**: atributo especial (aprendizado supervisionado), denominado rótulo ou classe
 - Classificação: discreto ou nominal $C_1, C_2, \dots, C_k$
 - Regressão: valores reais



Escala Nominal ou Categórica



- Valores são símbolos distintos que servem apenas para rotular ou identificar
 - Atributo “Sexo”: Masculino e Feminino
 - Atributo “Religião”: Católica, Protestante, Budismo,...
 - Atributo “Partido Político”: DEM, PT, PSDB, ...

- Não existem relações entre valores nominais - ordenação ou distância
 - Não faz sentido o teste “Masculino > Feminino”

- Apenas testes de igualdade podem ser feitos
 - “Sexo” = Masculino



- Os valores podem ser ordenados ou ranqueados
 - Toda subclasse pode ser comparada com uma outra em termos de uma relação da forma “maior que” ou “menor que”
 - Atributo “Temperatura”: Quente > Morno > Frio (no entanto, não faz sentido “Quente + Frio” ou “2*Morno”)
 - Distinção entre Nominal e Ordinal não é sempre clara
 - Atributo “Tempo”: Ensolarado, Nublado, Chuvoso

- Quantidades intervalares além de ordenadas, também possuem unidades constantes de medidas
 - Diferenças entre quaisquer dois pontos adjacentes em qualquer parte da escala são iguais
 - O ponto zero é arbitrário
 - Soma e produto não fazem sentido
 - As escalas intervalares mais familiares são as escalas de temperatura Fahrenheit e Celsius
 - Cada uma tem um ponto zero arbitrário e nenhum indica uma quantia nula ou ausência de temperatura
 - Podemos dizer que 80°F tem o dobro de temperatura de 40°F ?

- Difere da escala intervalar por possuir um zero absoluto
 - Todas as operações matemáticas são possíveis com medidas em escala de razão
 - Números reais
 - Atributo “Distância”: a distância entre um objeto e ele mesmo é zero
 - Atributo “Peso”: os aparelhos usados para medir peso têm um ponto zero absoluto

Para que tipos específicos de atributos?

- Compreender os diferentes tipos de escalas de medidas é importante por duas razões
 - O pesquisador deve identificar a escala de medida de cada atributo usado, de forma que dados não-métricos não sejam incorretamente usados como dados métricos e vice-versa
 - “Partido Político” > PSB não faz sentido, enquanto que
“Temperatura” > Frio ou
“Peso” < 38 fazem
 - A escala de medida é crítica ao determinar que algoritmos de aprendizado de máquina são mais apropriados

Conceitos e Definições

- Conjunto de exemplos: um conjunto de exemplos é composto por exemplos contendo valores de atributos bem como a classe associada.

Tempo	Temperatura	Humidade	vento	Joga
Sol	85	85	Não	Não
Sol	80	90	Sim	Não
Nublado	83	86	Não	Sim
Chuva	70	96	Não	Sim
Chuva	68	80	Não	Sim
Chuva	65	70	Sim	Não
Nublado	64	65	Sim	Sim
Sol	72	95	Não	Não
Sol	69	70	Não	Sim
Chuva	75	80	Não	Sim
Sol	75	70	Sim	Sim
Nublado	72	90	Sim	Sim
Nublado	81	75	Não	Sim
Chuva	71	91	Sim	Não

Conceitos e Definições



- **Classificador ou Hipótese:** dado um conjunto de exemplos de treinamento, um indutor gera como saída um classificador (também denominado hipótese ou descrição de conceito) de forma que, dado um novo exemplo, ele possa predizer com a maior precisão possível sua classe.
- **Ruído:** é comum no mundo real, trabalhar com dados imperfeitos. Eles podem ser derivados do próprio processo que gerou os dados, do processo de aquisição de dados, do processo de transformação ou mesmo devido a classes rotuladas incorretamente (por exemplo, exemplos com os mesmos valores de atributos mas com classes diferentes).



Conceitos e Definições



- **Missing Values (Valores Perdidos):** em geral, indicados por valores fora do escopo
 - **Tipos:** desconhecidos, não registrados, irrelevantes
 - Razões
 - Mau-funcionamento do equipamento
 - Mudanças na definição do experimento
 - Incapacidade de medição
- Valores perdidos podem, de fato, significar alguma coisa
 - A maioria dos métodos de aprendizado não assumem isto
 - No entanto, este tipo de informação pode ser codificado como um valor adicional



Conceitos e Definições

- **Modo de aprendizado:**
 - **Não incremental (batch):** sempre que todo o conjunto de treinamento deva estar presente para o aprendizado.
 - **Incremental:** o indutor apenas tenta atualizar a hipótese antiga sempre que novos exemplos são adicionados ao conjunto de treinamento.
- Taxa de Erro de um classificador h :
 - Compara a classe verdadeira de cada exemplo com o rótulo atribuído pelo classificador induzido.

$$\text{err}(h) = \frac{1}{n} \sum_{i=1}^n \| y_i \neq h(x_i) \|$$

- Precisão do Classificador: complemento da taxa de erro denotado por $\text{acc}(h)$

$$\text{acc}(h) = 1 - \text{err}(h)$$

- Erro de Regressão: calcula-se a distância entre o valor real com o atribuído pela hipótese induzida. Pode-se usar: erro médio quadrado ou a distância absoluta média.

$$\text{mse-err}(h) = \frac{1}{n} \sum_{i=1}^n (y_i - h(x_i))^2$$

$$\text{mad-err}(h) = \frac{1}{n} \sum_{i=1}^n |y_i - h(x_i)|$$

Resources: Datasets



- UCI Repository: <http://www.ics.uci.edu/~mlearn/MLRepository.html>
- UCI KDD Archive: <http://kdd.ics.uci.edu/summary.data.application.html>
- Statlib: <http://lib.stat.cmu.edu/>
- Delve: <http://www.cs.utoronto.ca/~delve/>

