Data

BUSINESS ANALYST

CHANGE AGENT

Role

Improves business process as intermediary between business and IT

Languages

SQL

Mindset

Resilient project juggler

Skills & Talents

- Basic tools (e.g. MS Office)
- Data visualization tools (e.g. Tableau)
- Conscious listening and storytelling
- Business Intelligence understanding
- Data modeling

DATA ANALYST

DATA DETECTIVE

Role

Collects, processes and performs statistical data analyses

Languages

R, Python, HTML, Javascript, C/C++, SQL

Mindset

Intutive data junkie with high "figure-it-out" quotient

Skills & Talents

- Spreadsheet tools (e.g. Excel)
- Database systems (SQL and NO SQL based)
- Communication & visualization
- Math, Stats, Machine Learning

Who is a Data Scientist?

Clean and Transform the Data

STUDENT LOAN APPLICATION

Personal Information

(Last)

(First)

(Middle Initial)

Home Telephone
( ) -

Other Telephone
( ) -

(City)

ip)

Student Loan Data

Loan ID#	Graduated	Loan Type	Loan Balance	Next Payment	Months Delinquent	Defaulted
10148975	Yes	1	31567	327	0	N
19773966	Yes	3	27909	-	3	N
25220947	Yes	2	11,463	243	25	Y
17090812	No	2	29801	255	15	Y
23956341	Yes	3	18755	173	0	N
12680900	Yes	1	16,211	122	7	N
23435111	No	1	5064	84	0	N

One-Hot Encoding

Student Loan Data

Graduated	Loan 1	Loan 2	Loan 3	Loan Balance	Next Payment	Months Delinquent	Defaulted
1	1	0	0	31567	327	0	N
1	0	0	1	27909	200	3	N
1	0	1	0	11463	243	25	Y
0	0	1	0	29801	255	15	Y
1	0	0	1	18755	173	0	N
1	1	0	0	16211	122	7	N
0	1	0	0	5064	84	0	N

DataFrame

Student Loan Data

Graduated	Loan Balance	Next Payment	Months Delinquent	Defaulted
1	31567	327	0	0
1	27909	200	3	0
1	11463	243	25	1
0	29801	255	15	1
1	18755	173	0	0
1	16211	122	7	0
0	5064	84	0	0
0	17198	154	0	0
1	21309	201	0	0
0	14693	193	4	1

Student Loan Data

1	31567	327	0	0
1	27909	200	3	0
1	11463	243	25	1
0	29801	255	15	1
1	18755	173	0	0
1	16211	122	7	0
0	5064	84	0	0
0	17198	154	0	0
1	21309	201	0	0
0	14693	193	4	1

Common Machine Learning Algorithms

Linear Regression

Logistic Regression

Support Vector Machine

Decision Tree

$$\hat{f}(X)$$

The Prediction

$$\overset{\text{output}}{\hat{y}} = \hat{f}(\overset{\text{input}}{X})$$

Linear Regression

equation of a line

$$y = mx + b$$

equation of a line

$$y = mx + b$$

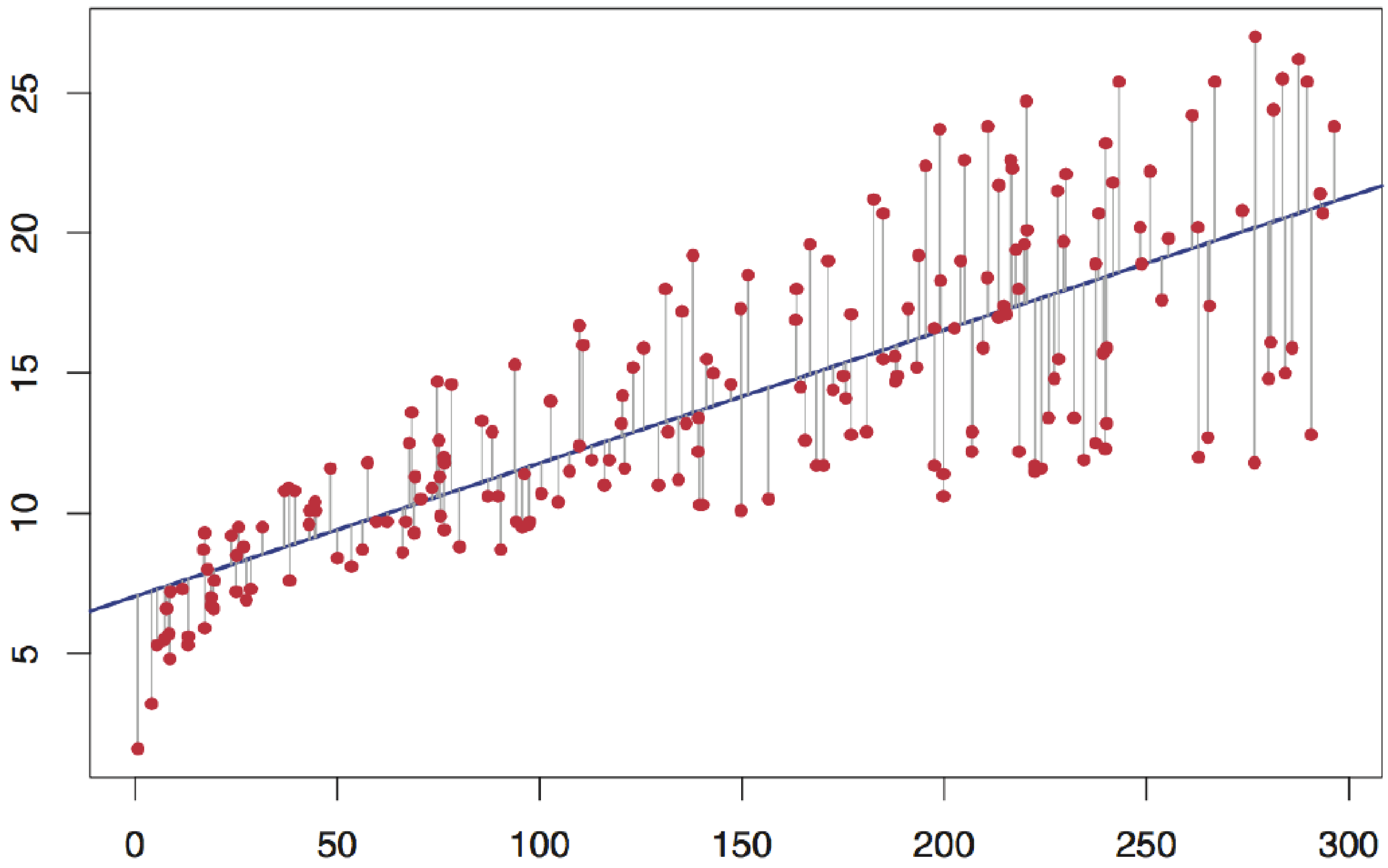
linear regression

$$y = \beta_0 + \beta_1 x$$

Simple Linear Regression

Input	learned coefficients (weights)	output
x_1	β_0 , β_1	y

$$y = \beta_0 + \beta_1 x$$


25
20
15
10
5
0
50
100
150
200
250
300

Weekly Hours Spent Studying

Gender	Number of Classes	Social Accounts	Hours Studying
Male	2	1	7.5
Male	4	3	12.25
Female	4	3	12.75
Female	3	4	7.75
Female	4	2	14
Male	2	3	5.75
Female	5	1	18.25

Multiple Linear Regression

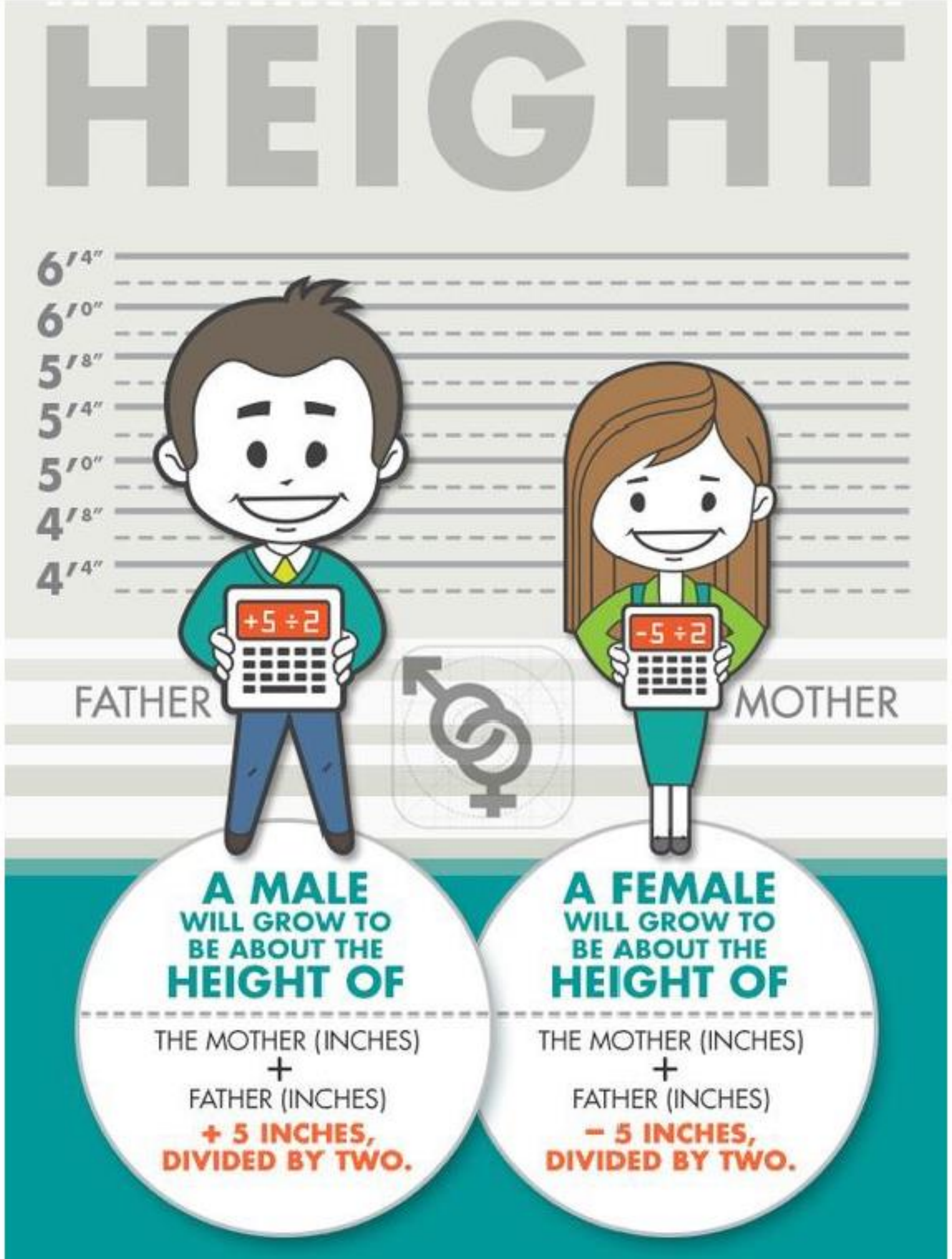
$$\text{Hrs. Studying} = 1.63 + 3.51x_1 + .27x_2 - 1.08x_3$$

x_1 = number of classes

x_2 = gender (m=0, female=1)

x_3 = number of social accounts

HEIGHT

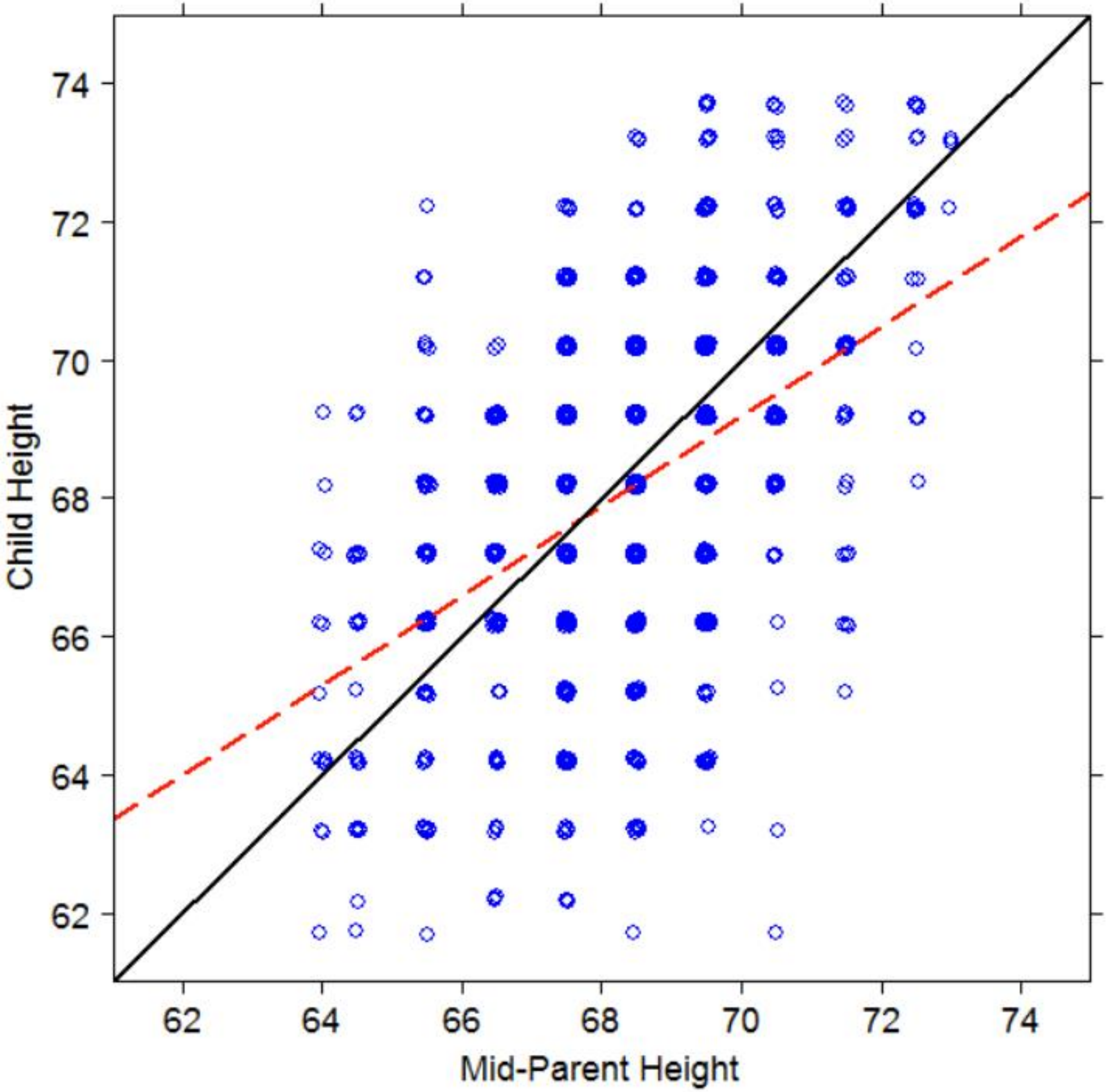


A MALE WILL GROW TO BE ABOUT THE **HEIGHT OF**

THE MOTHER (INCHES)
+
FATHER (INCHES)
+ 5 INCHES, DIVIDED BY TWO.

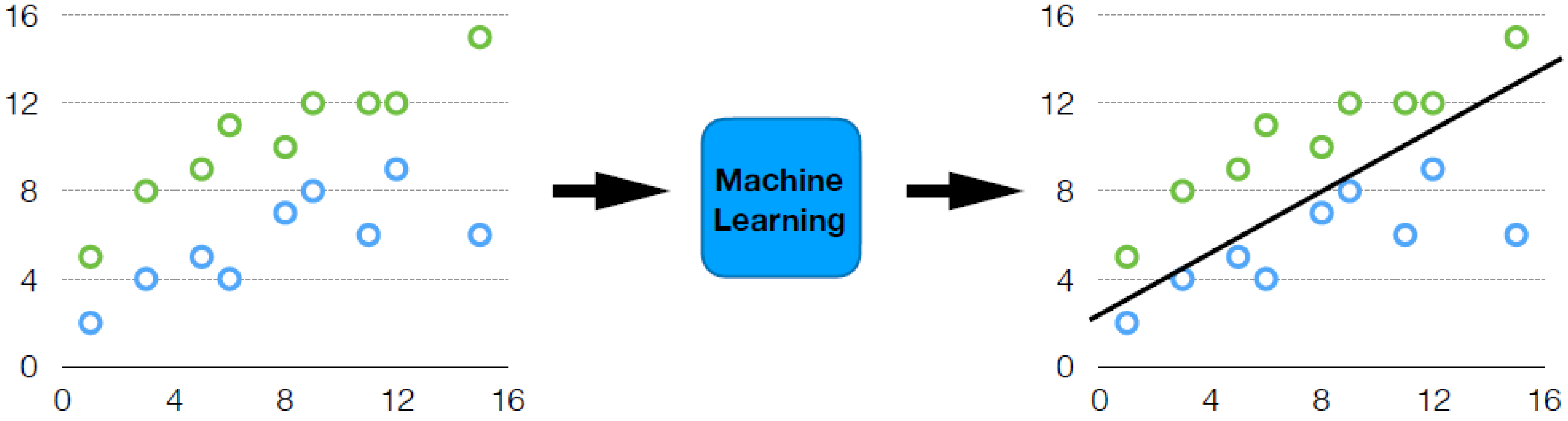
A FEMALE WILL GROW TO BE ABOUT THE **HEIGHT OF**

THE MOTHER (INCHES)
+
FATHER (INCHES)
− 5 INCHES, DIVIDED BY TWO.


Child Height
Mid-Parent Height
62
64
66
68
70
72
74
62
64
66
68
70
72
74

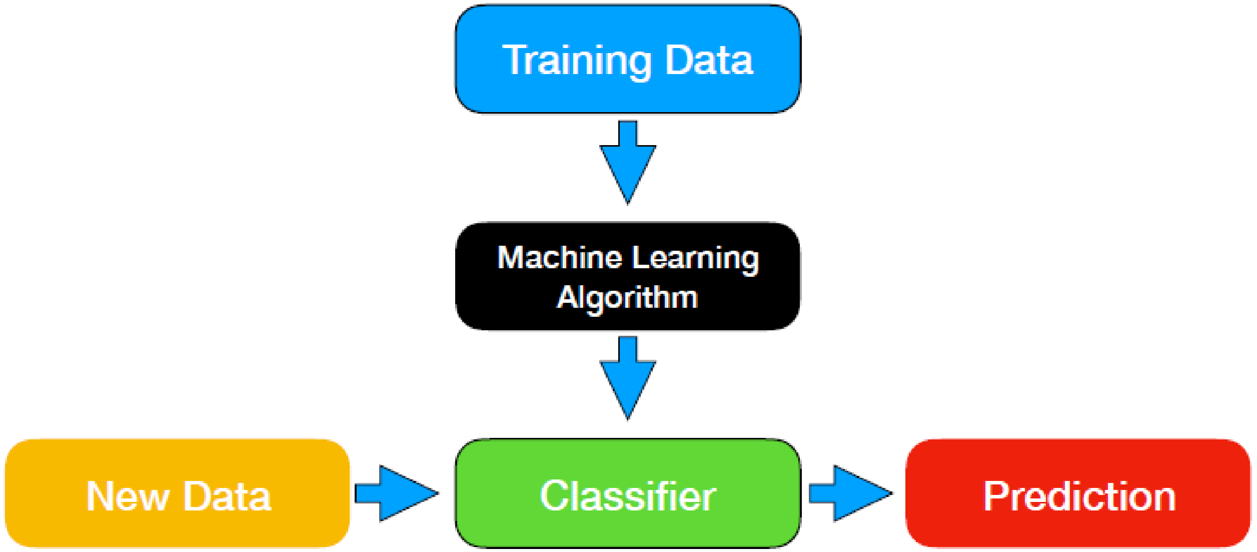
CLASSIFIER

What is a classifier?



Machine Learning

Learning Process



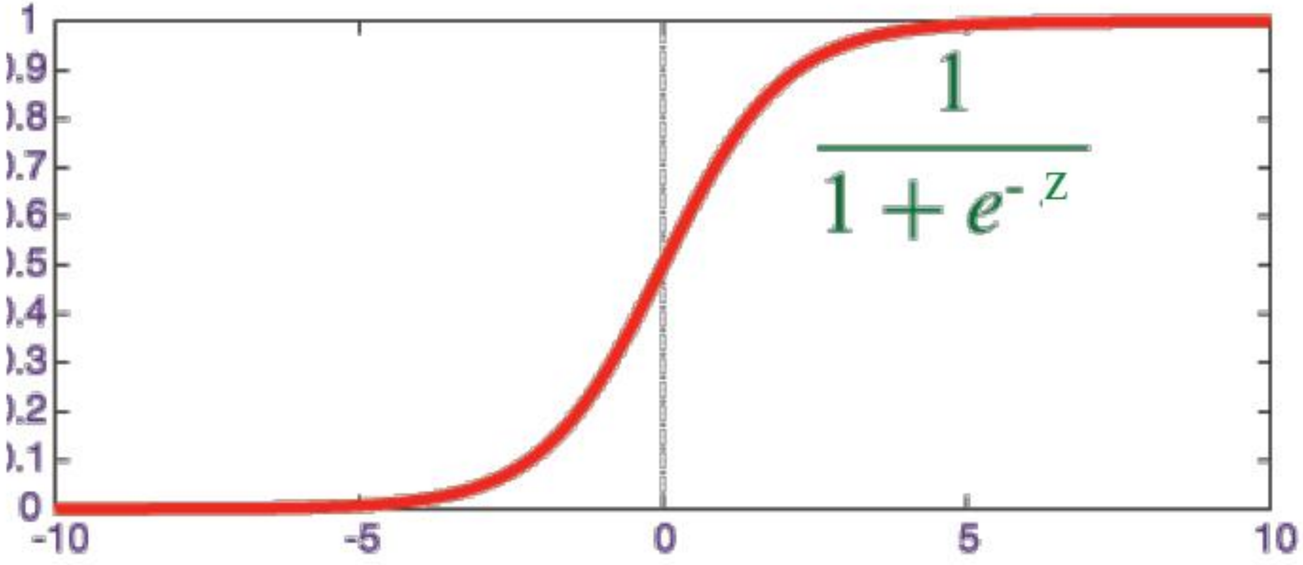
Logistic Regression

Logistic Regression

intermediate step

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Logistic (Sigmoid) Function

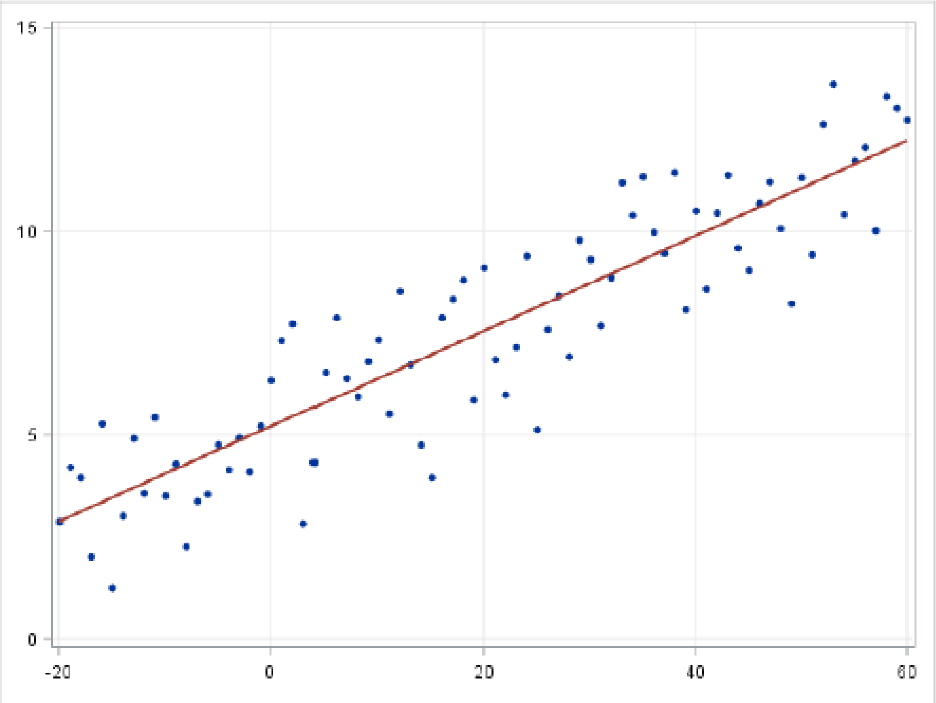


Z

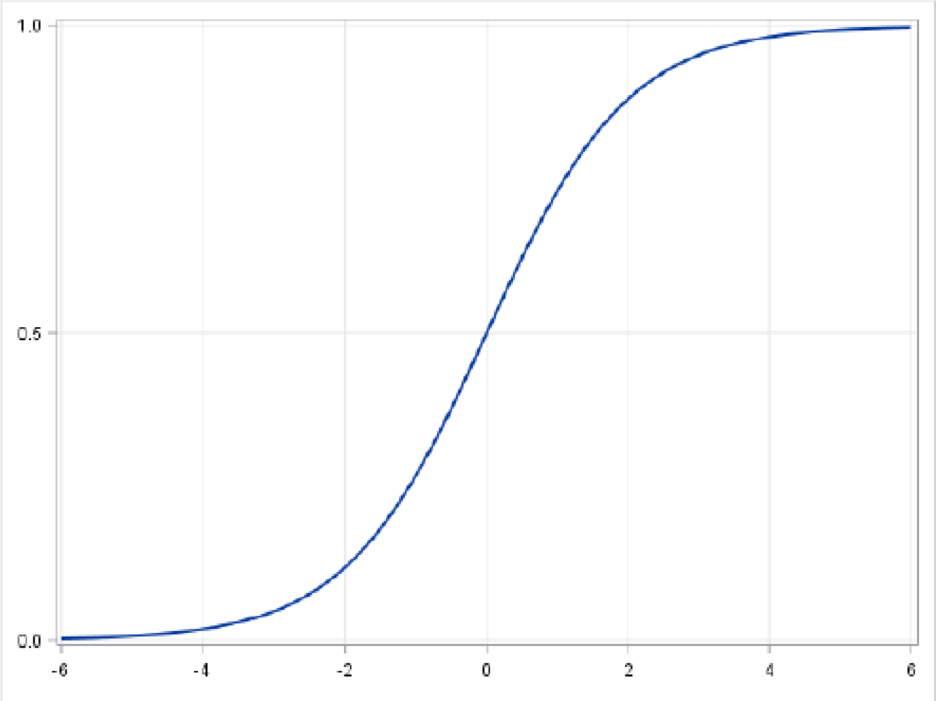
Logistic Regression

$$\hat{f}(X) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2)}}$$

Linear Regression



Logistic Regression

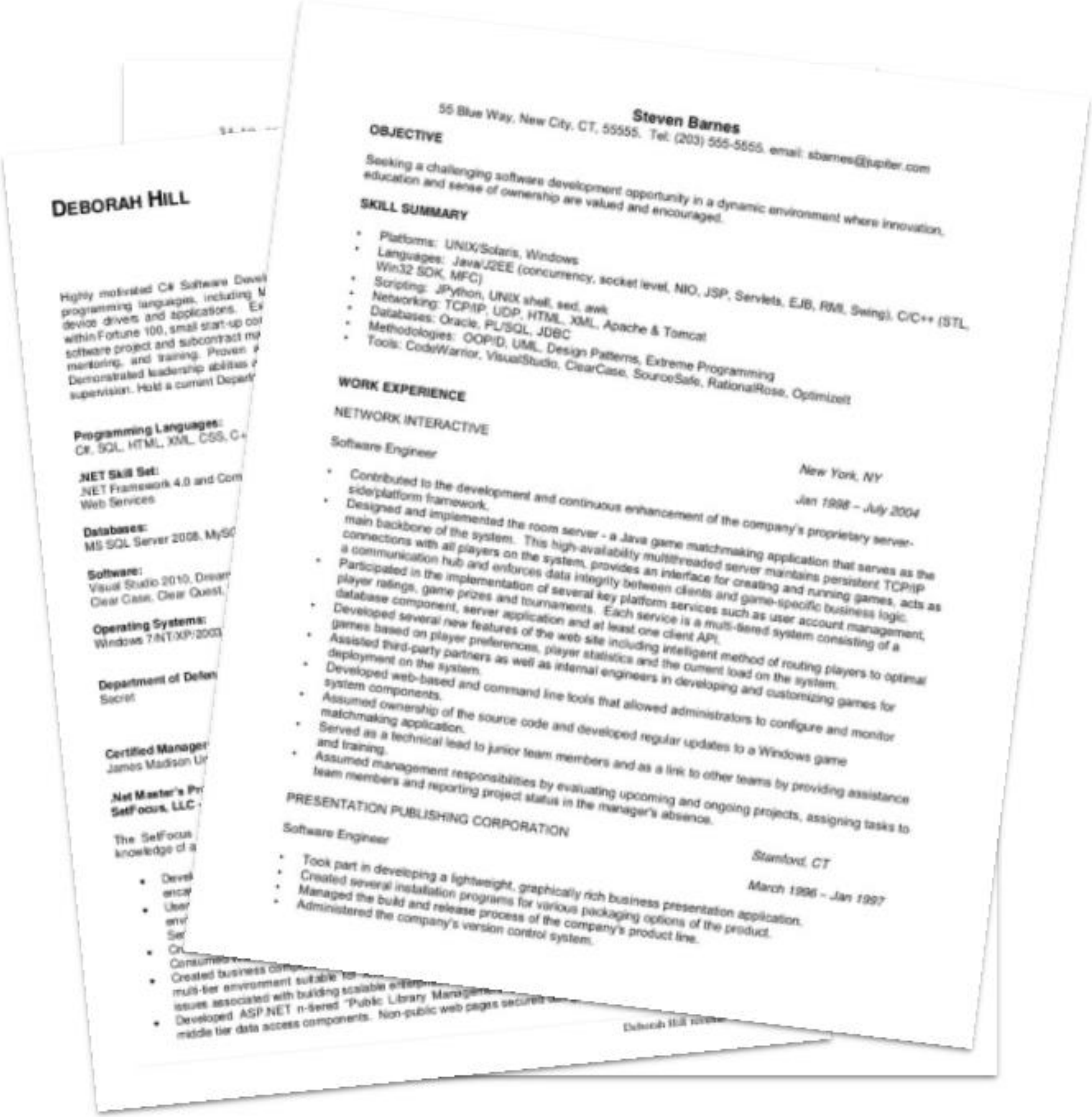


positive class = 1

negative class = 0

if probability >= 0.5 : predict 1

if probability < 0.5 : predict 0



Steven Barnes

55 Blue Way, New City, CT, 55555. Tel: (203) 555-5555. email: sbarnes@jupiter.com

OBJECTIVE

Seeking a challenging software development opportunity in a dynamic environment where innovation, education and sense of ownership are valued and encouraged.

SKILL SUMMARY

- Platforms: UNIX/Solaris, Windows
- Languages: Java/J2EE (concurrency, socket level, NIO, JSP, Servlets, EJB, RMI, Swing), C/C++ (STL, Win32 SDK, MFC)
- Scripting: JPython, UNIX shell, sed, awk
- Networking: TCP/IP, UDP, HTML, XML, Apache & Tomcat
- Databases: Oracle, PL/SQL, JDBC
- Methodologies: OOP/D, UML, Design Patterns, Extreme Programming
- Tools: CodeWarrior, VisualStudio, ClearCase, SourceSafe, RationalRose, OptimizeIt

WORK EXPERIENCE

NETWORK INTERACTIVE

Software Engineer

New York, NY

Jan 1998 – July 2004

- Contributed to the development and continuous enhancement of the company's proprietary server-side/platform framework.
- Designed and implemented the room server - a Java game matchmaking application that serves as the main backbone of the system. This high-availability multithreaded server maintains persistent TCP/IP connections with all players on the system, provides an interface for creating and running games, acts as a communication hub and enforces data integrity between clients and game-specific business logic.
- Participated in the implementation of several key platform services such as user account management, player ratings, game prizes and tournaments. Each service is a multi-tiered system consisting of a database component, server application and at least one client API.
- Developed several new features of the web site including intelligent method of routing players to optimal games based on player preferences, player statistics and the current load on the system.
- Assisted third-party partners as well as internal engineers in developing and customizing games for deployment on the system.
- Developed web-based and command line tools that allowed administrators to configure and monitor system components.
- Assumed ownership of the source code and developed regular updates to a Windows game matchmaking application.
- Served as a technical lead to junior team members and as a link to other teams by providing assistance and training.
- Assumed management responsibilities by evaluating upcoming and ongoing projects, assigning tasks to team members and reporting project status in the manager's absence.

PRESENTATION PUBLISHING CORPORATION

Software Engineer

Stamford, CT

March 1996 – Jan 1997

- Took part in developing a lightweight, graphically rich business presentation application.
- Created several installation programs for various packaging options of the product.
- Managed the build and release process of the company's product line.
- Administered the company's version control system.

DEBORAH HILL

Highly motivated C# Software Devel programming languages, including M device drivers and applications. Ex within Fortune 100, small start-up co software project and subcontract ma mentoring, and training. Proven Demonstrated leadership abilities supervision. Hold a current Depart

Programming Languages:
C#, SQL, HTML, XML, CSS, C+

.NET Skill Set:
.NET Framework 4.0 and Com
Web Services

Databases:
MS SQL Server 2008, MySQ

Software:
Visual Studio 2010, Dream
Clear Case, Clear Quest,

Operating Systems:
Windows 7/NT/XP/2003

Department of Defen
Secret

Certified Manager
James Madison Ur

.Net Master's Pr
SetFocus, LLC

The SetFocus knowledge of a

- Devel enca
- User env Ser
- Cr Consume
- Created business comp multi-tier environment suitable for issues associated with building scalable enterp
- Developed ASP.NET n-tiered "Public Library Manage middle tier data access components. Non-public web pages secured

Deborah Hill

Gender	Years Exp.	Source 1	Source 2	Source 3	Phone Screen	On-site Interview
1	3	0	1	0	9	1
0	2	0	0	1	7.5	0
0	2	1	0	0	7	0
0	4	0	1	0	8.5	1
1	4	0	1	0	9.5	1
1	2	0	1	0	6.5	0
0	3	1	0	0	8	0
1	2	0	0	1	8	0
1	4	0	1	0	9	1
0	4	0	1	0	7	1

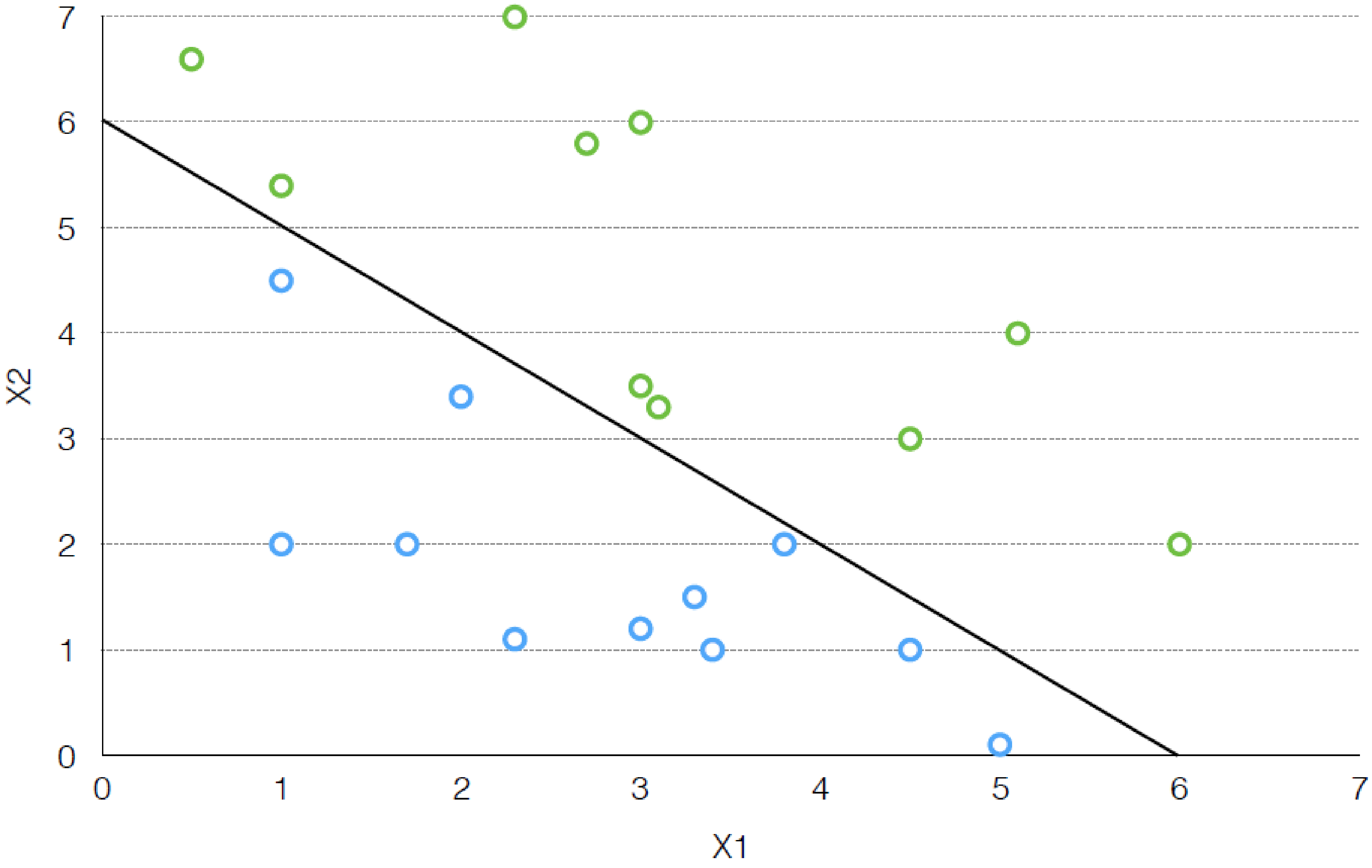
Decision Boundary

On-site Interview = $-6 + 1x_1 + 1x_2$

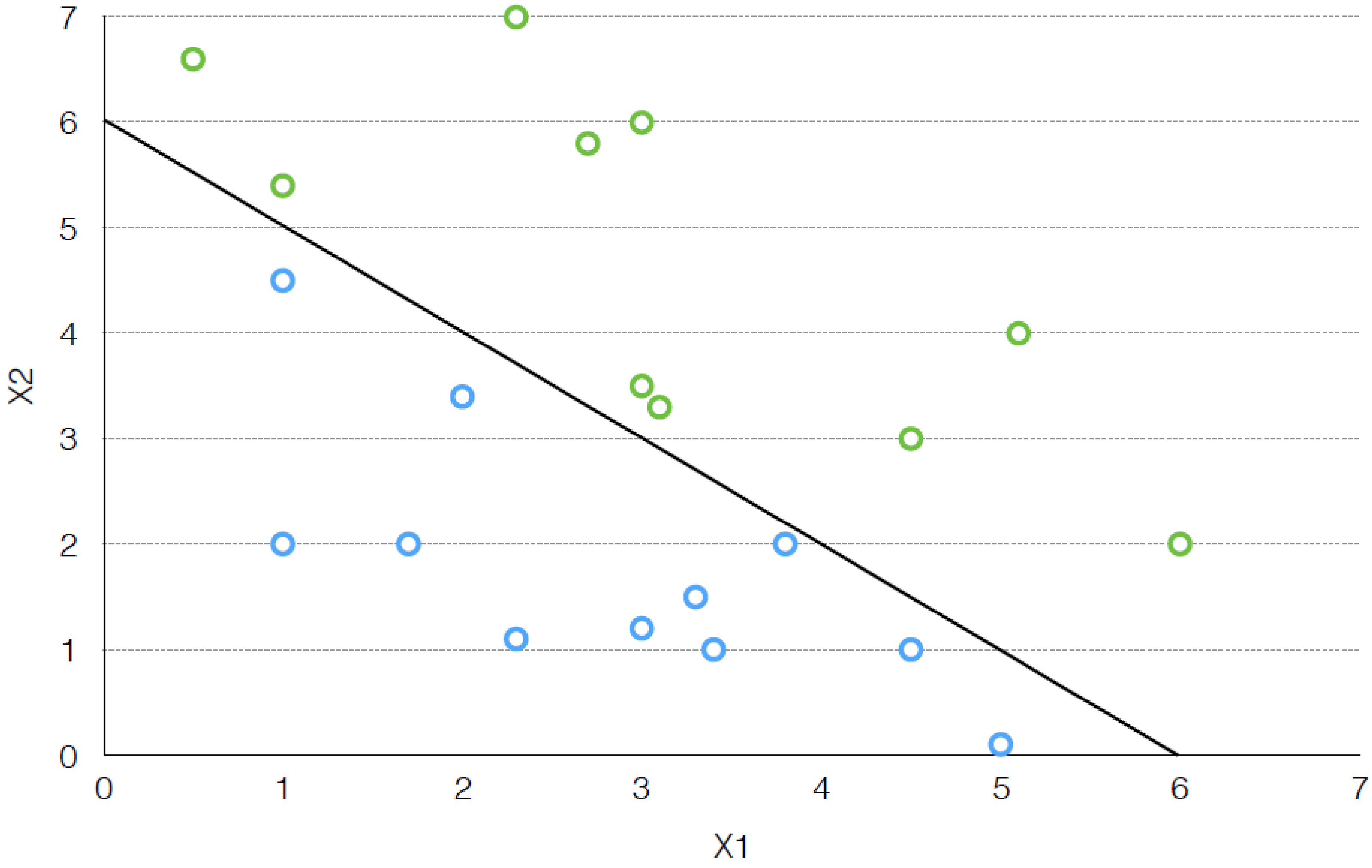
predict 1 when $-6 + 1x_1 + 1x_2 >= 0$

predict 0 when $-6 + 1x_1 + 1x_2 < 0$

Decision Boundary = $-6 + 1x_1 + 1x_2$



$$\hat{f}(X) = \frac{1}{1+e^{-(-6 + 1x_1 + 1x_2)}}$$



Support Vector Machine

Support Vector Machine

Large Margin Classifier

if $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 >= 1$: predict 1

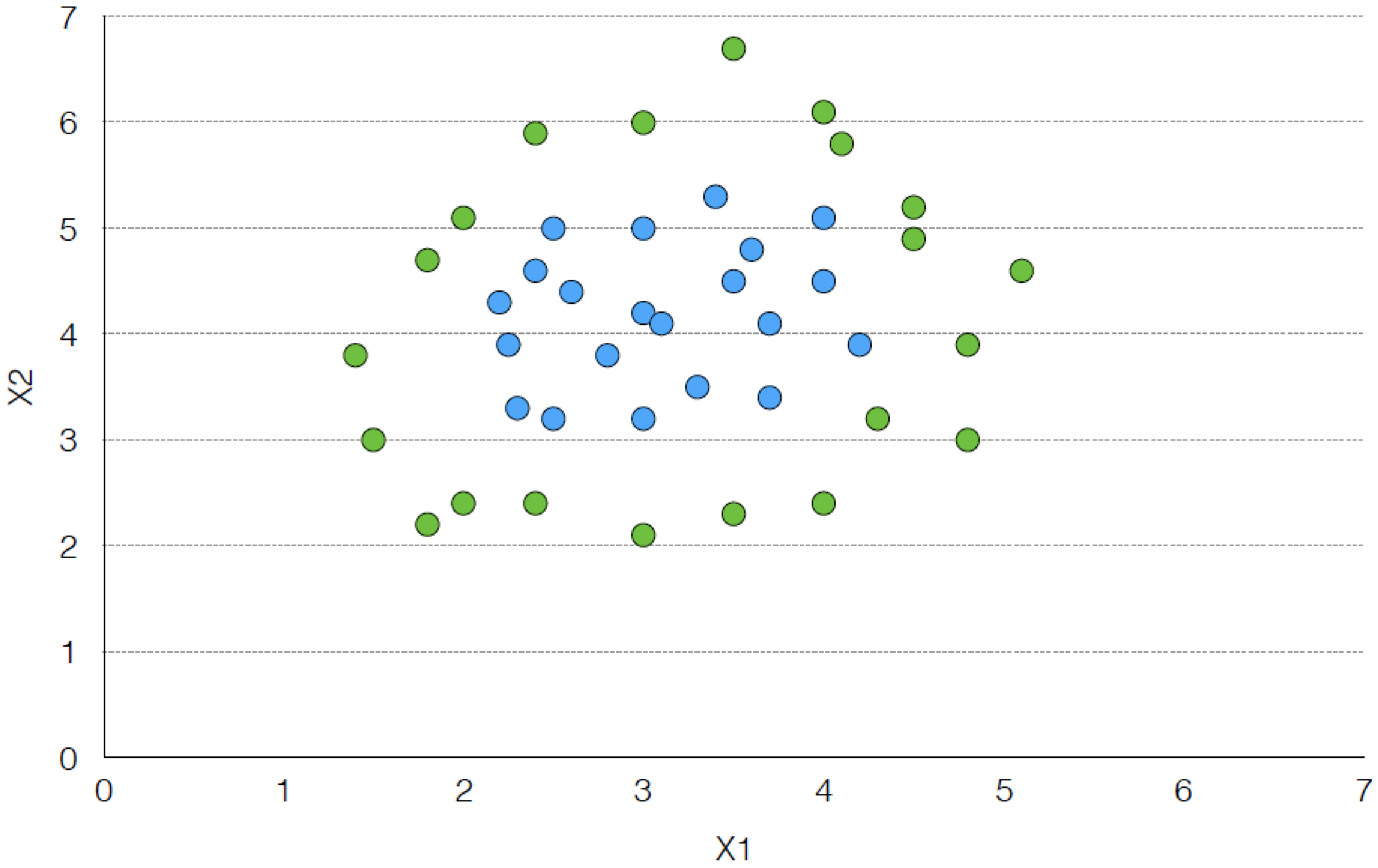
if $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 < -1$: predict 0

NFL
SCOUTING
COMBINE
2016

40-yard dash	Weight	Height	Drafted
5.10	290	74	1
4.92	275	75.5	1
4.43	178	69	0
4.62	221	74.5	1
4.91	248	75	0
5.53	303	77	0
4.47	189	71	1
4.56	205	71	1
4.75	267	73	0
4.84	261	74	1

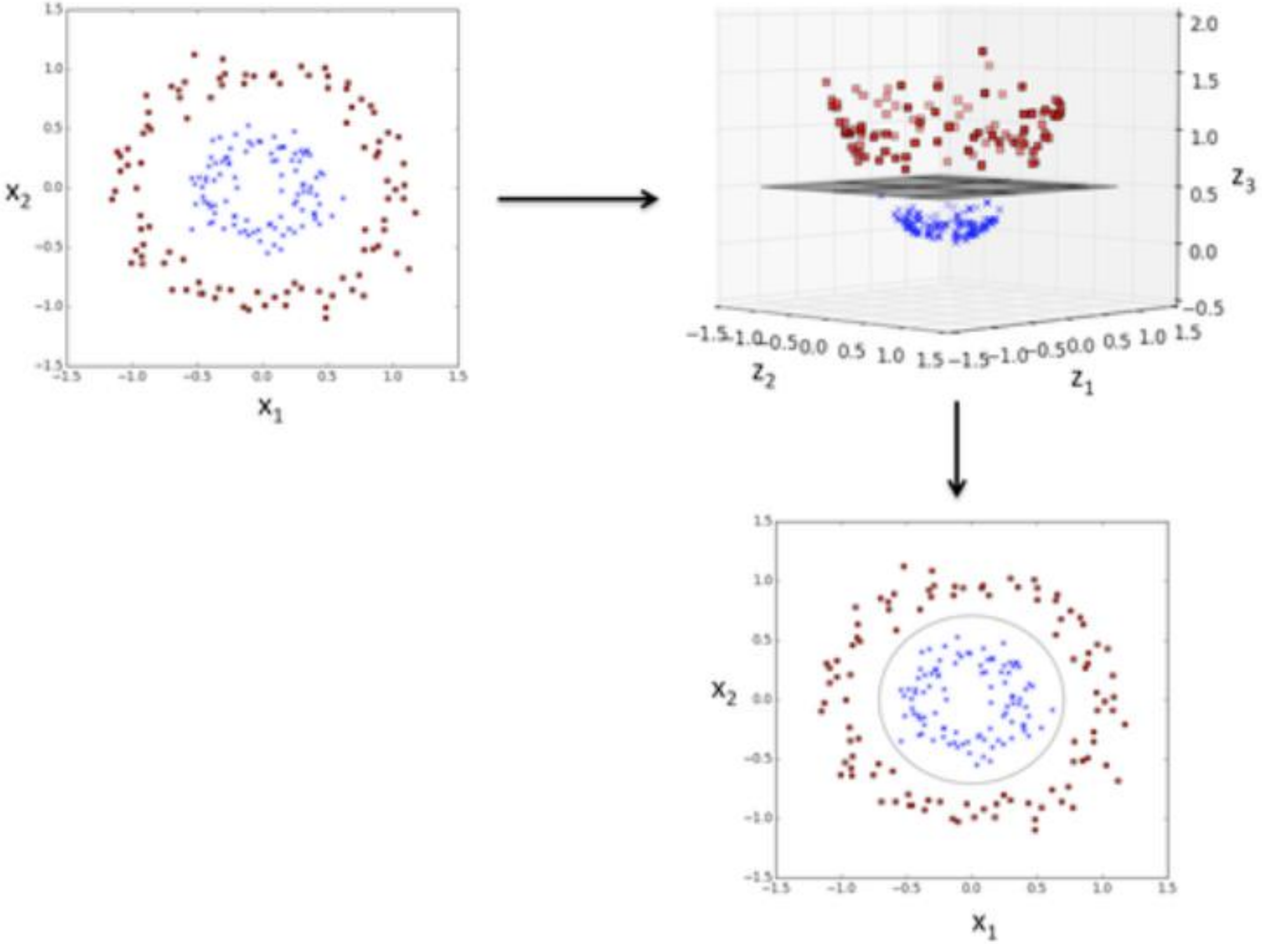
Feature Engineering

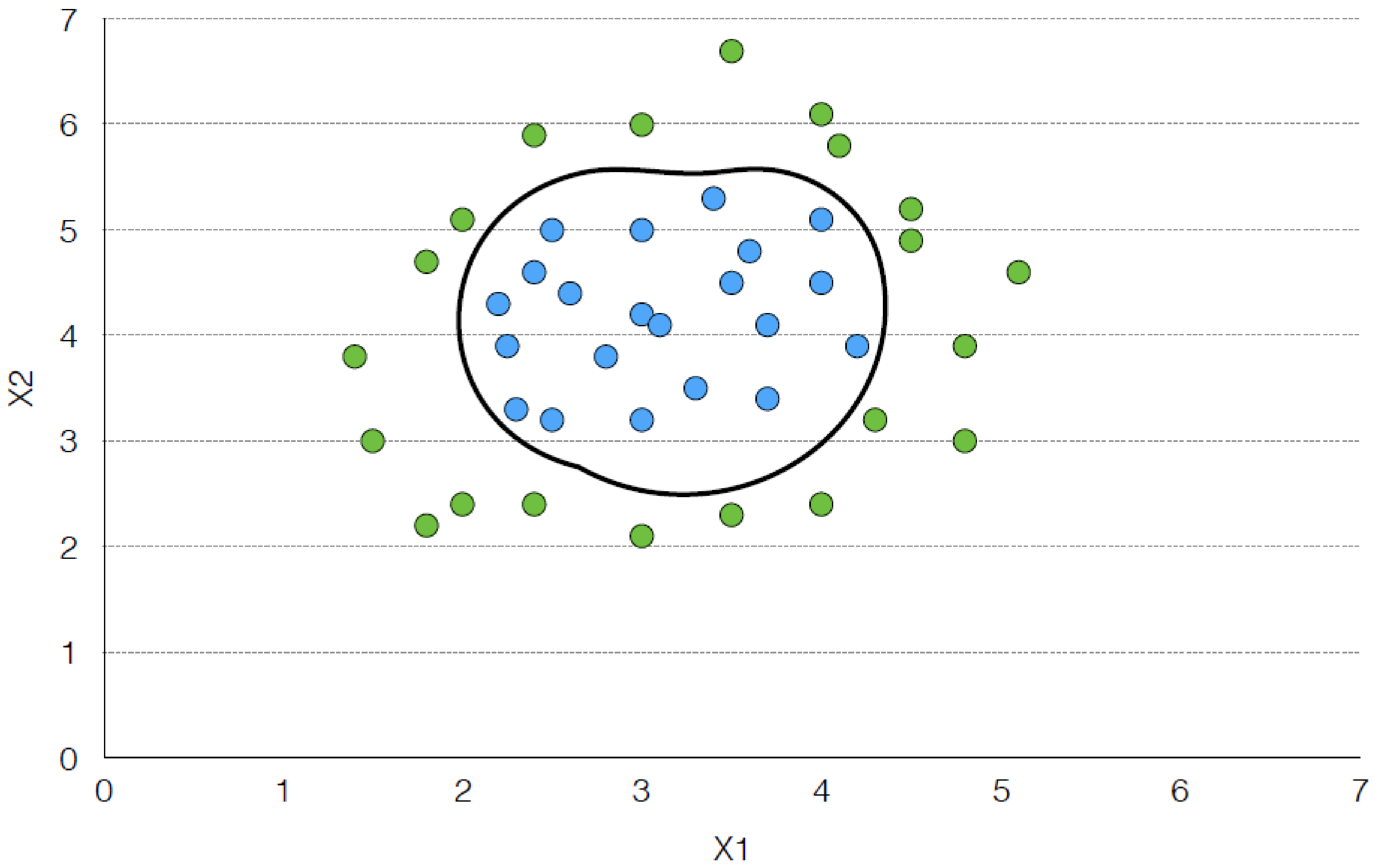
40-yard dash	BMI (wt/ht^2)	Drafted
5.10	37.2	1
4.92	33.9	1
4.43	26.3	0
4.62	28	1
4.91	31	0
5.53	35.9	0
4.47	26.4	1
4.56	28.6	1
4.75	35.2	0
4.84	33.5	1


X2
X1
0
1
2
3
4
5
6
7
0
1
2
3
4
5
6
7

Kernel

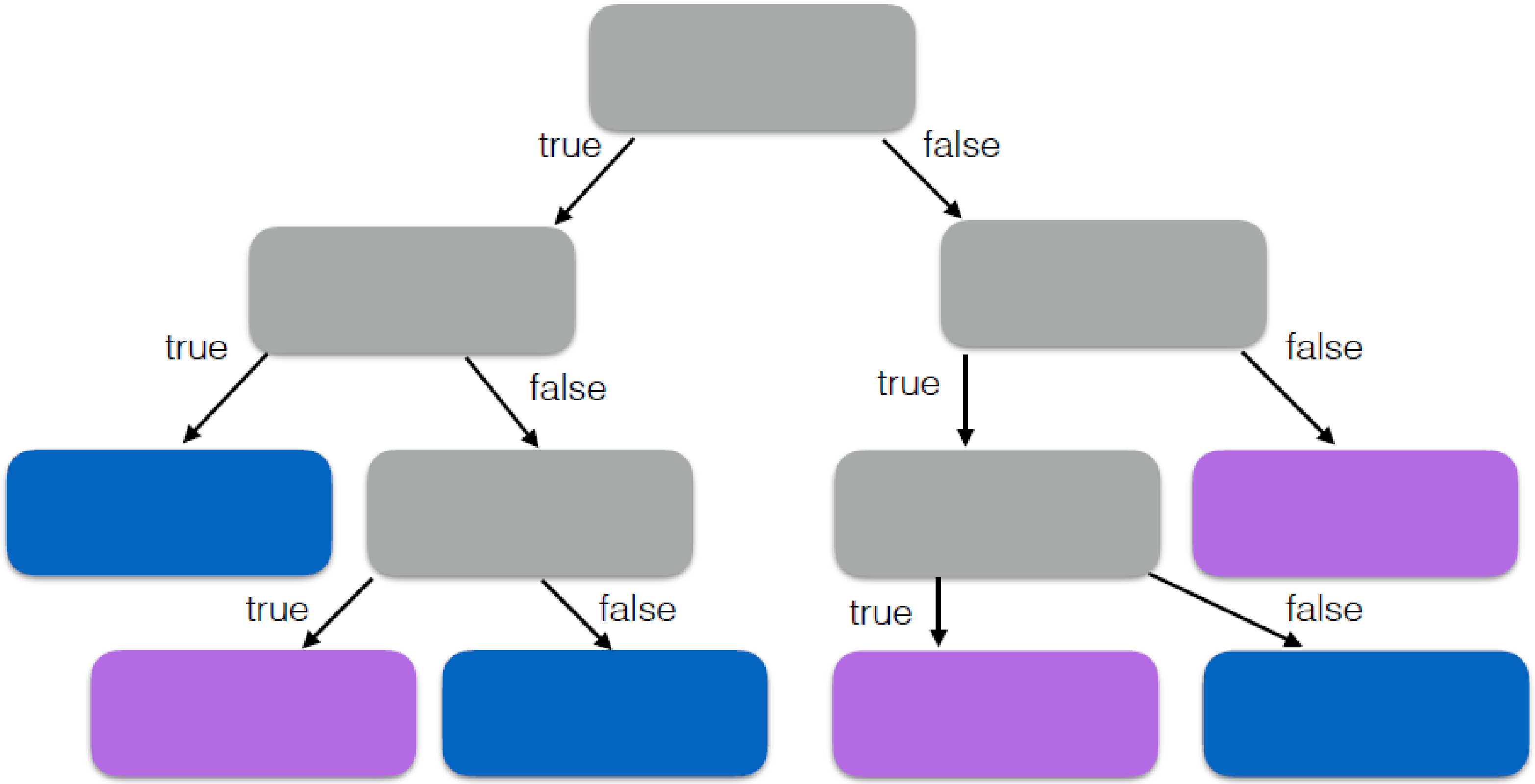
non-linear classification


x_2
x_1
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
z_3
2.0
1.5
1.0
0.5
0.0
−0.5
z_2
z_1
x_2
x_1


7
6
5
4
3
2
1
0
X2
0
1
2
3
4
5
6
7
X1

Decision Tree

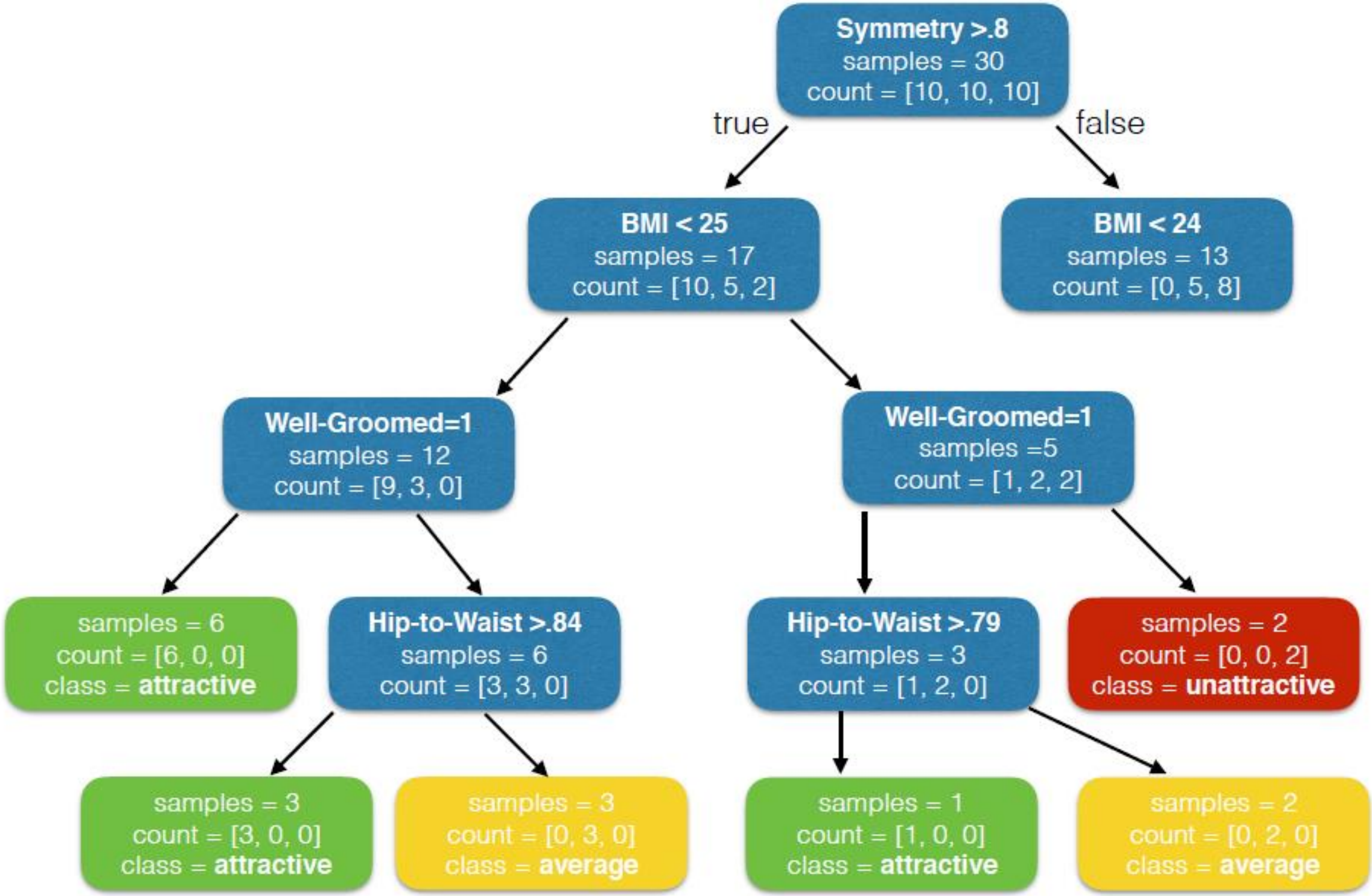
Decision Tree



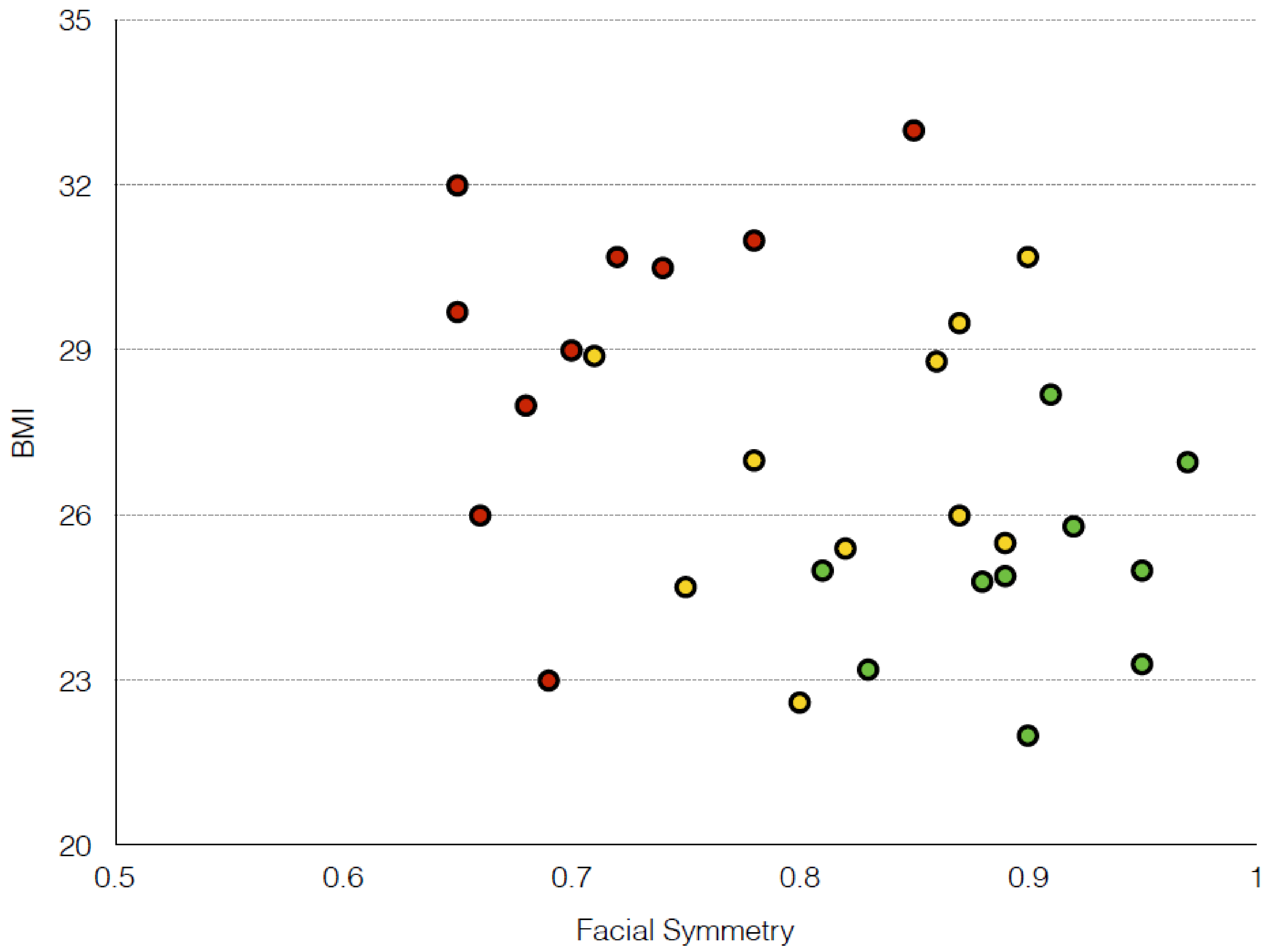
Short-term Attractiveness

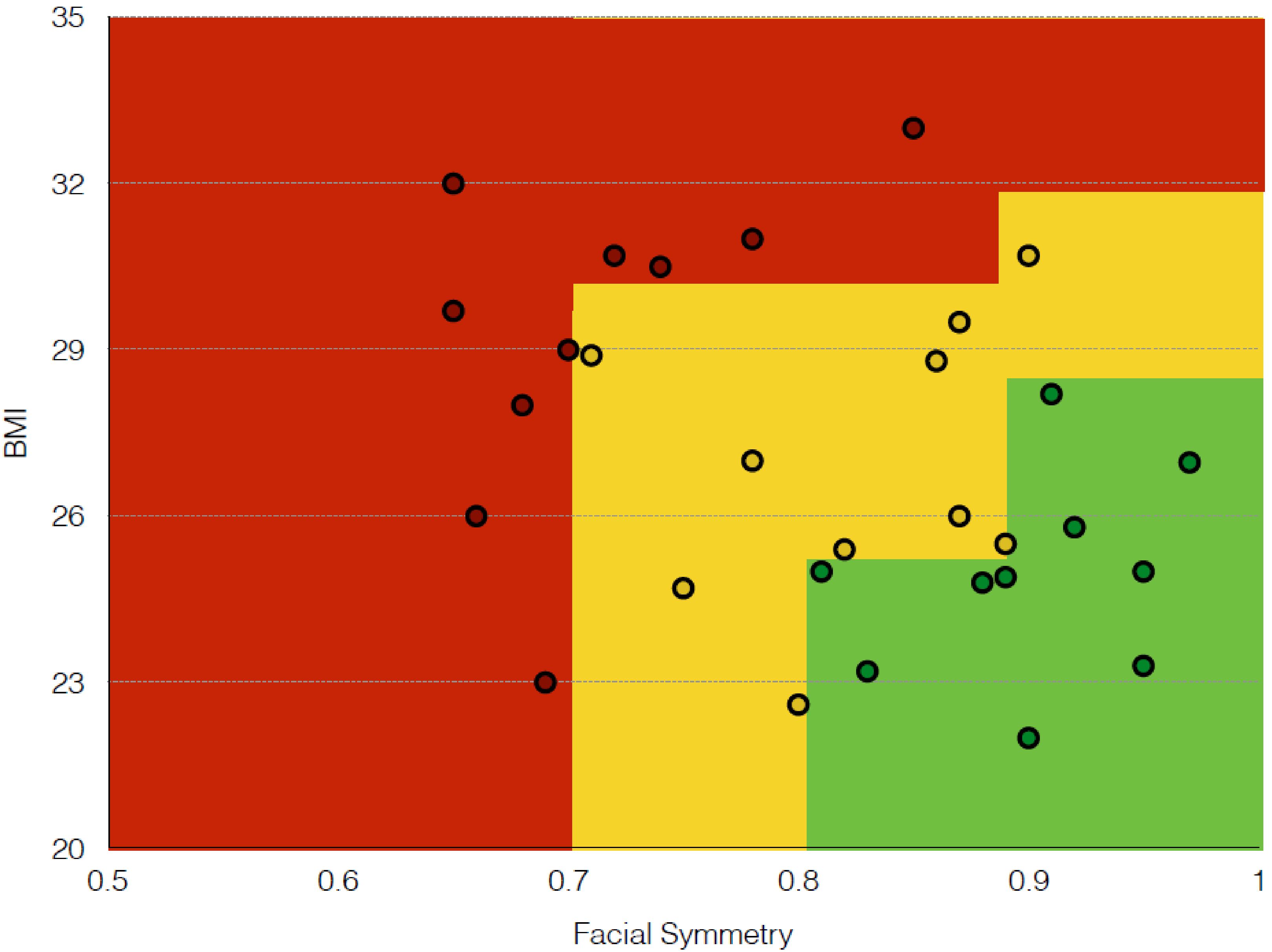
Short-term Attractiveness

Facial Symmetry	BMI	Waist-to-Hip	Well-Groomed
0.9	23.4	0.93	1
0.85	27.9	0.87	0
0.65	27.1	0.79	1
0.85	22.6	0.91	1
0.9	30.3	0.82	0
0.75	29.0	0.82	0
0.85	22.3	0.89	1
0.7	37.6	0.73	0
0.85	24.2	0.85	0


Symmetry >.8
samples = 30
count = [10, 10, 10]
true
false
BMI < 25
samples = 17
count = [10, 5, 2]
BMI < 24
samples = 13
count = [0, 5, 8]
Well-Groomed=1
samples = 12
count = [9, 3, 0]
Well-Groomed=1
samples =5
count = [1, 2, 2]
samples = 6
count = [6, 0, 0]
class = attractive
Hip-to-Waist >.84
samples = 6
count = [3, 3, 0]
Hip-to-Waist >.79
samples = 3
count = [1, 2, 0]
samples = 2
count = [0, 0, 2]
class = unattractive
samples = 3
count = [3, 0, 0]
class = attractive
samples = 3
count = [0, 3, 0]
class = average
samples = 1
count = [1, 0, 0]
class = attractive
samples = 2
count = [0, 2, 0]
class = average

[att, ave, un]

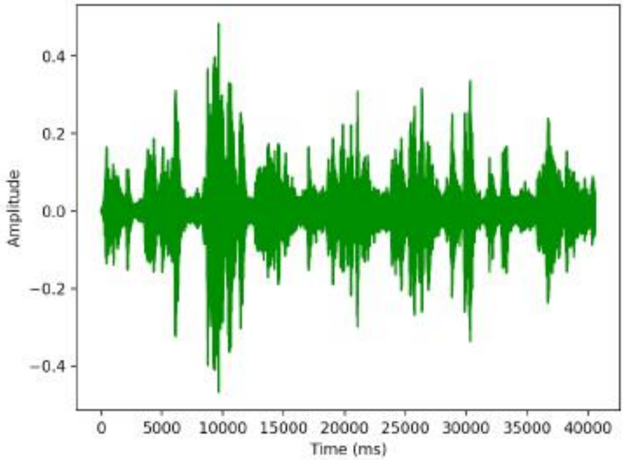

35
32
29
26
23
20
BMI
0.5
0.6
0.7
0.8
0.9
1
Facial Symmetry


35
32
29
26
23
20
BMI
0.5
0.6
0.7
0.8
0.9
1
Facial Symmetry

Structured Data

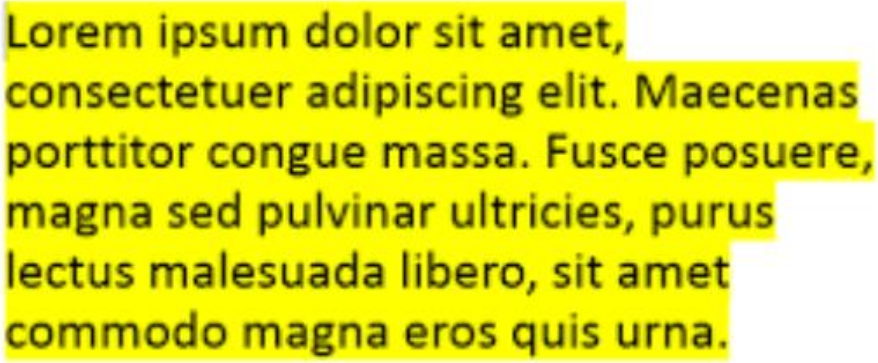
Age	Weight	Gender	BMI	Diabetes
47	192	M	23.4	No
53	164	F	27.2	Yes
68	214	M	25.2	Yes
43	151	F	24.8	No

Unstructured Data



Audio

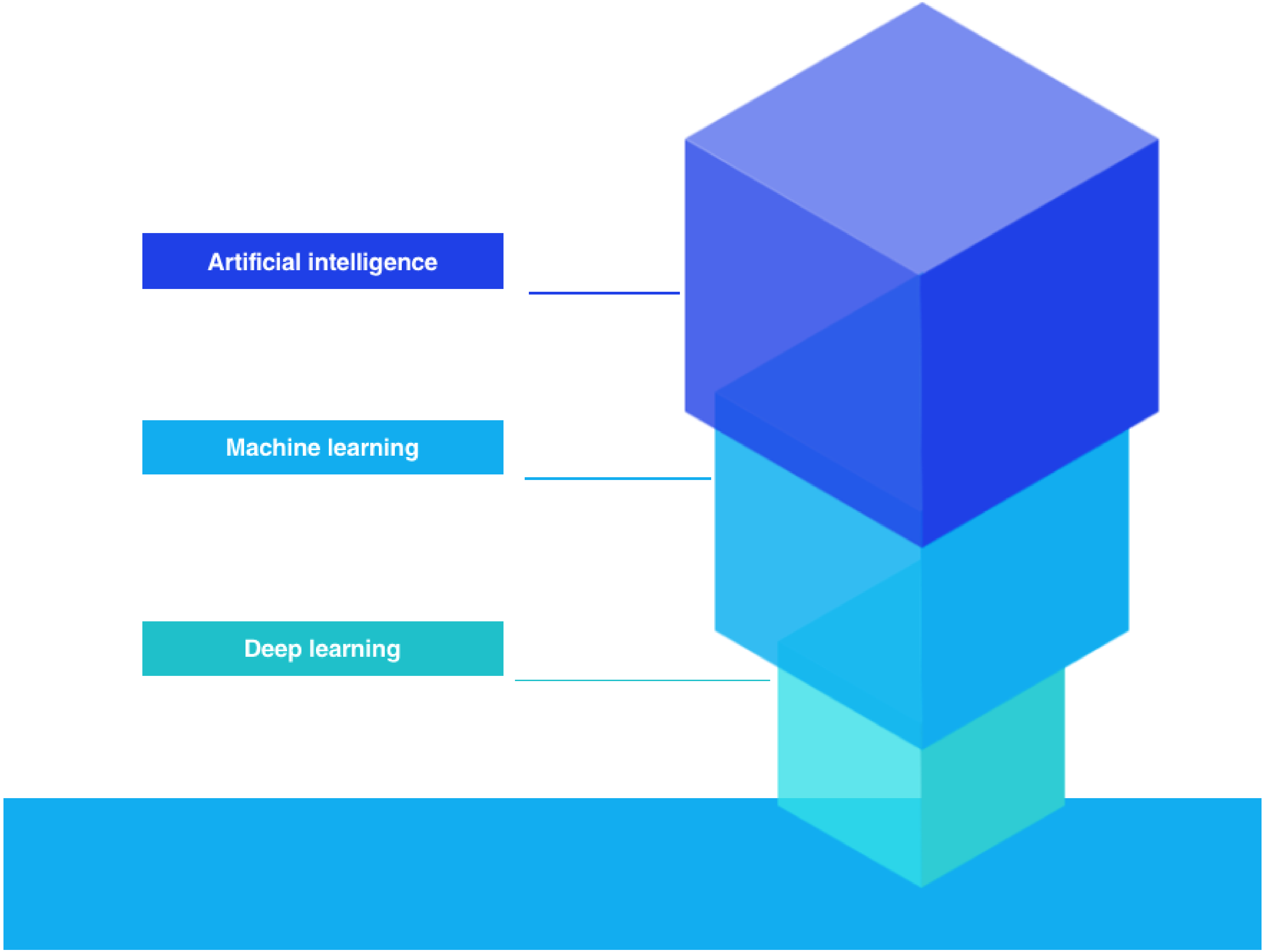
Image



Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.

Text

Deep Learning



Artificial intelligence

Machine learning

Deep learning

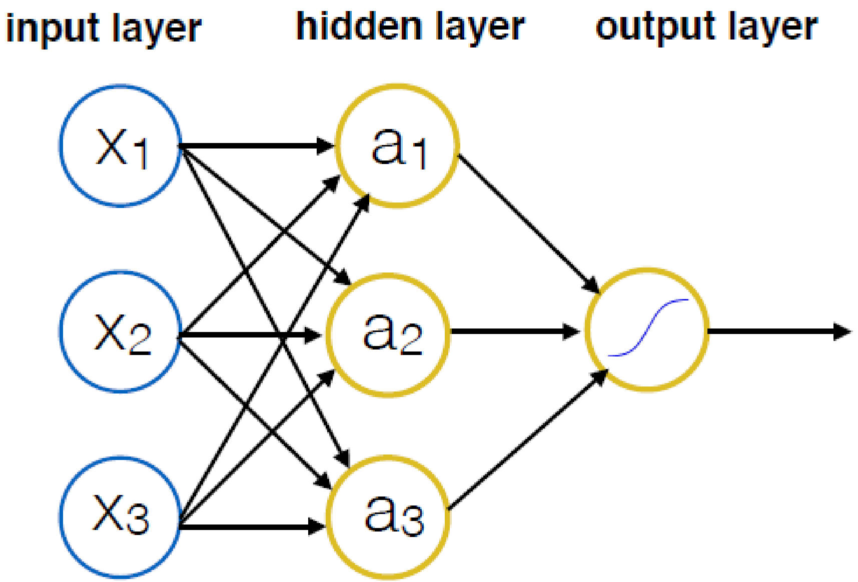
Siri

Spotify

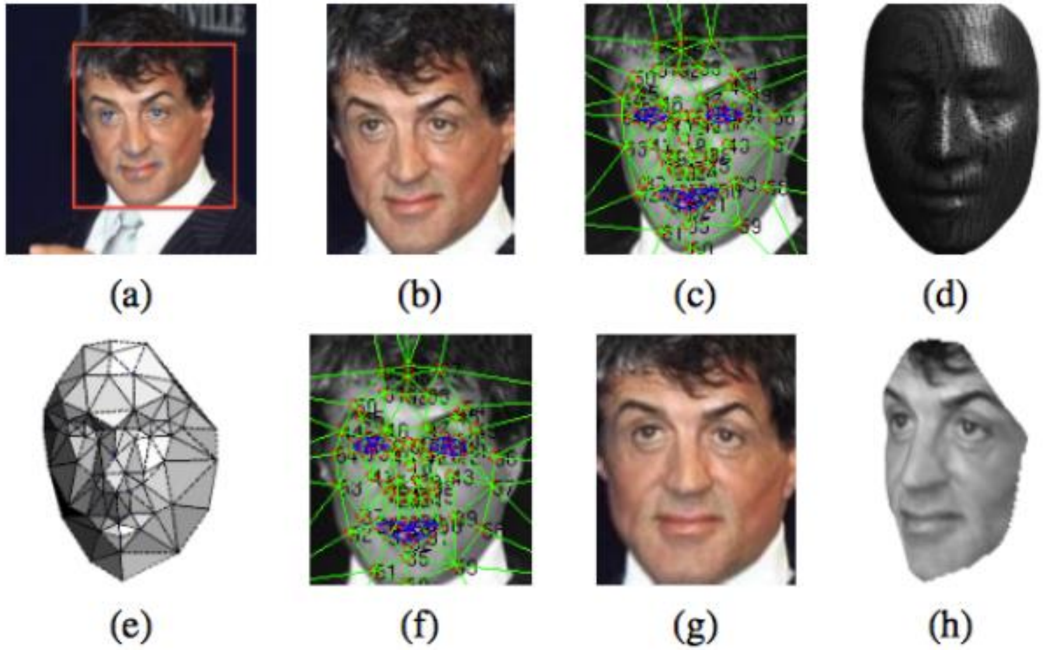
Google Photos

Neural Network

(feed forward network)



Facebook - DeepFace



Meetup