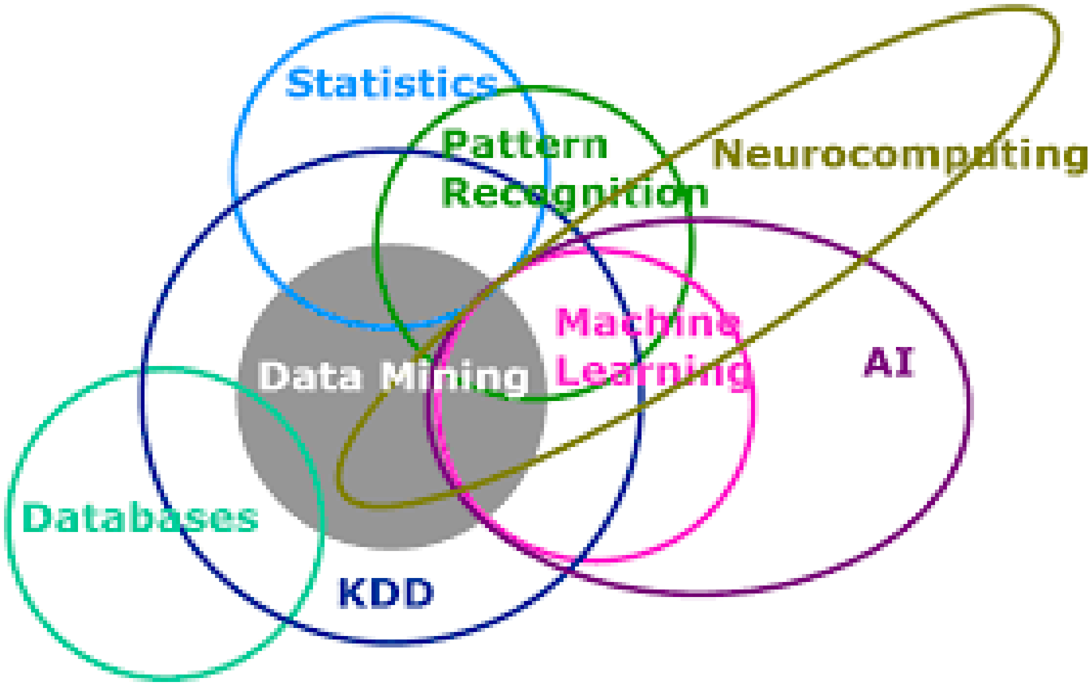
In the name of Allah the most Beneficial ever merciful

بِسۡمِ ٱللَّهِ ٱلرَّحۡمَٰنِ ٱلرَّحِيمِ



Artificial Intelligence (AI) in Software Engineering

Naive Bayes, Perceptron Rule Exam Discussion

Copyright © 2021, Dr. Humera Tariq

Department of Computer Science , Univeristy of Karachi (DCS-UBIT)
8th June 2021

Overview of Classification System

2nd Last week

Feature Selection
Feature Normalization
Feature Representation
K-Means Classifier

Step 1: Problem Statement or Research Question

1. **RQ1:** Can deep learning models such as CNNs offer competitive performance on software requirements classification?

2. **RQ2:** Can leveraging the power of Big Data when vectorizing our documents with pre-trained word embeddings boost CNN performance on software requirements classification?

Software Requirements Classification Using Word Embeddings and Convolutional Neural Networks (calpoly.edu)

Automatic Extraction of

Design Decision Relationships

from a Task Management System

RuppelMatthias_Thesis.pdf

UNIVERSITY OF KARACHI

Step 2: Background Study of Requirements

Table 2.1: Examples of functional and non-functional requirements from the NFR dataset.

Requirement Type	*Requirements Text*
Functional	"The system will notify affected parties when changes occur affecting classes including but not limited to class cancellations class section detail changes and changes to class offerings for a given quarter."
Performance	"Any interface between a user and the automated system shall have a maximum response time of 5 seconds unless noted by an exception below."
Scalability	"The product shall be capable of handling up to 1000 concurrent requests. This number will increase to 2000 by Release 2. The concurrency capacity must be able to handle peak scheduling times such as early morning and late afternoon hours."
Security	"User access should be limited to the permissions granted to their role(s) Each level in the PCG hierarchy will be assigned a role and users will be assigned to these roles. Access to functionality within RFS system is dependent on the privileges/permission assigned to the role."

ITY OF KARACHI

Step 3: Foundation of Classification

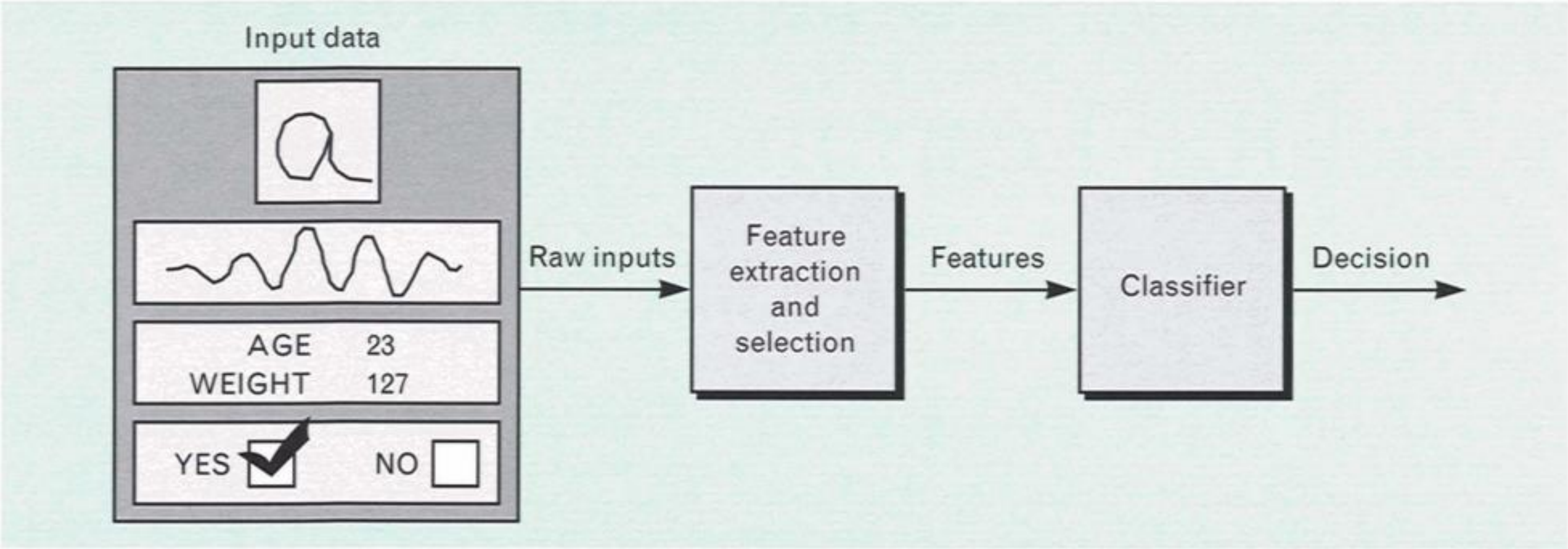
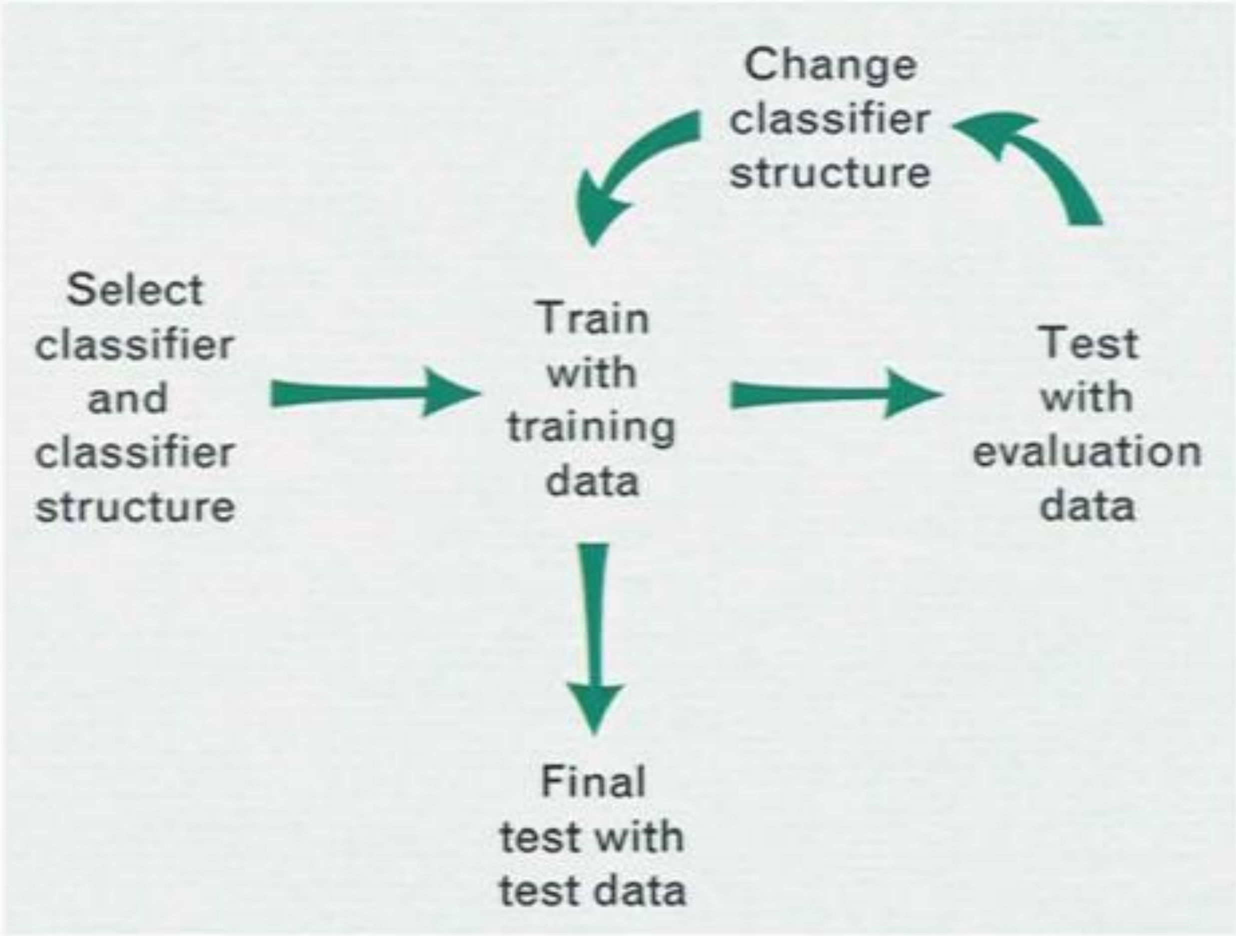


FIGURE 1. A simple pattern-classification system with image, waveform, categorical, and binary inputs.

[LNKnet: Neural network, machine-learning, and statistical software for pattern classification (mit.edu)]

UNIVERSITY OF KARACHI

Step 4: Foundation of Experiments



UNIVERSITY OF KARACHI

Step 3: Background Study of Text and Document Classification

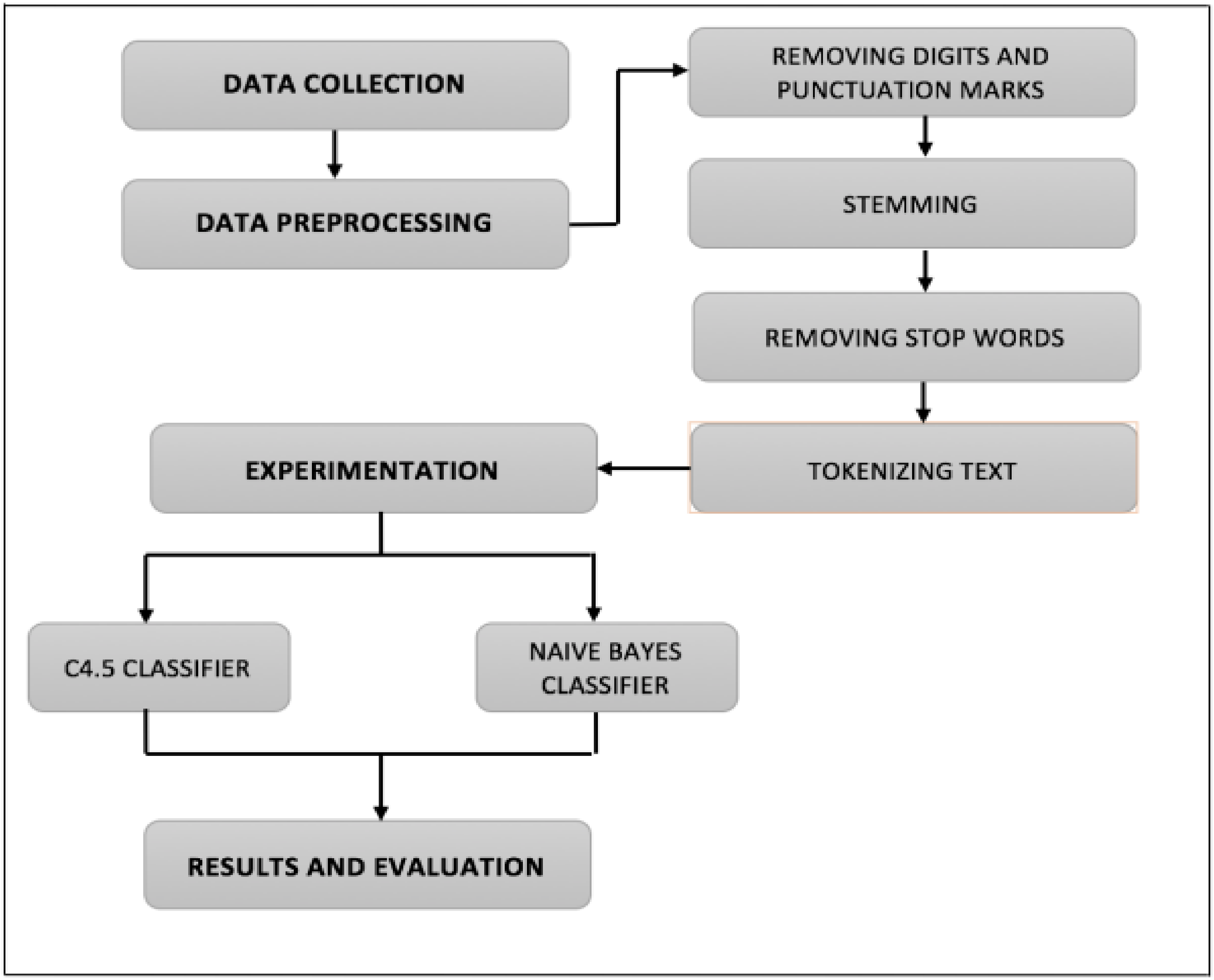
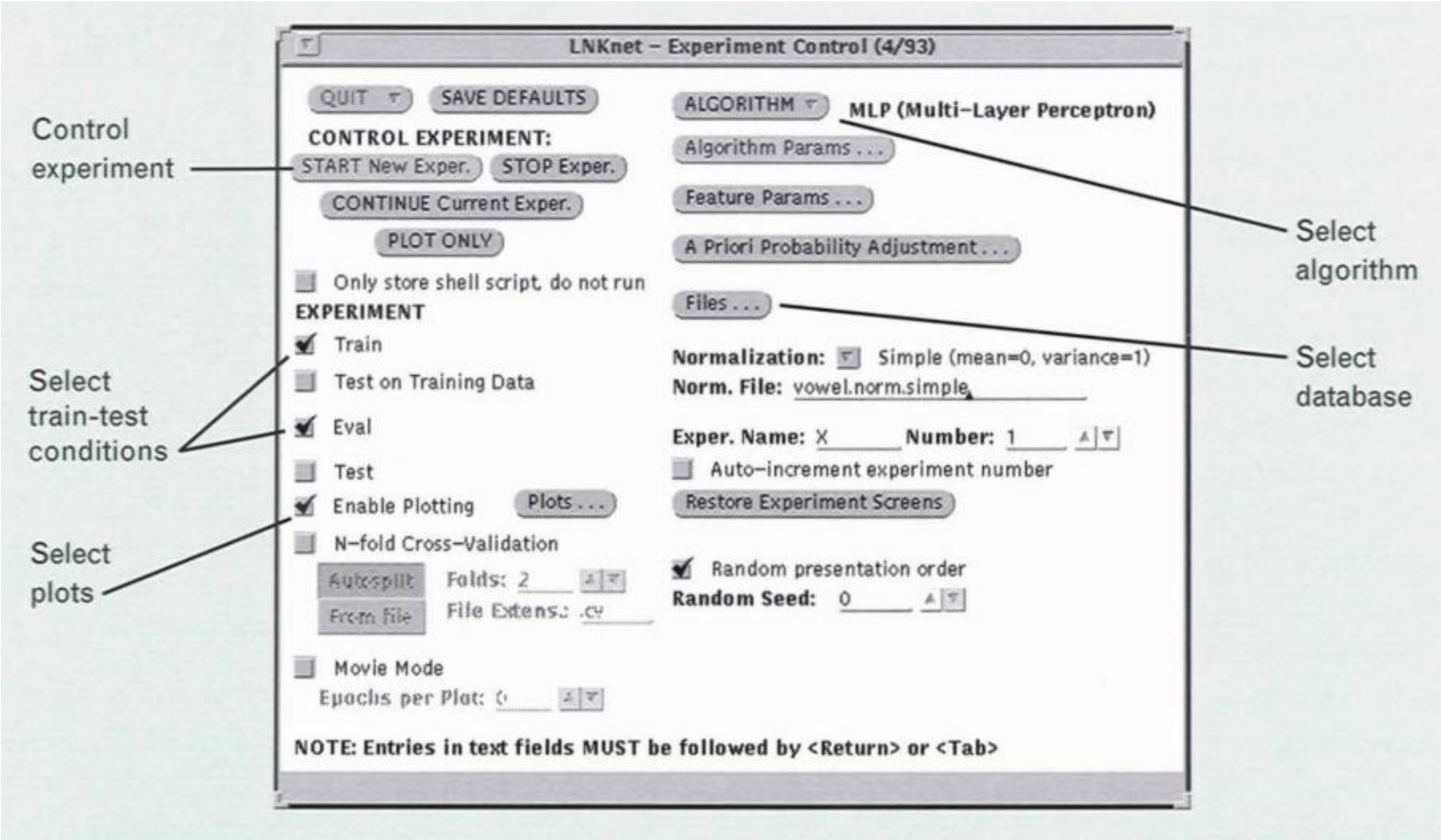


Figure 1 Overall methodology

UNIVERSITY OF KARACHI

Step 5: ML Website / Desktop Software



Practice Exercise 1
Normalization

Normalization

✓ The fact that input variables now have unit variance is an example of feature normalisation, which is a prerequisite for many ML algorithms.

Standardisation (Z-score Normalization)	Max-Min Normalization
$x_{\text{stand}} = \dfrac{x - \text{mean}(x)}{\text{standard deviation } (x)}$	$x_{\text{norm}} = \dfrac{x - \min(x)}{\max(x) - \min(x)}$

input	standardized	normalized
0.0	-1.336306	0.0
1.0	-0.801784	0.2
2.0	-0.267261	0.4
3.0	0.267261	0.6
4.0	0.801784	0.8
5.0	1.336306	1.0

Show all working steps for any three rows of given Table.

UNIVERSITY OF KARACHI

Practice Exercise 2

Assign Labels -> Supervised Learning

2

Few Abbreviation used at backend for Network Attachment/Configuration

CPN	Customer Premises Network
(ePurse)	Common Electronic Purse
(GPS)	Global Platform Specification
CEP	Currency Exchange Platform
CNG	Customer Network Gateways (User Equipment)
PCG	Public consulting Group (Privacy Policy)

UNIVERSITY OF KARACHI

2 Label given requirement with ‘Yes’ or ‘No’

SRS	*Requirements Text*	*Security-Related?*
ePurse	“All load transactions are on-line transactions. Authorization of funds for load transactions must require a form of cardholder verification. The load device must support on-line encrypted PIN or off-line PIN verification”	Y
	“A single currency cannot occupy more than one slot. The CEP card must not permit a slot to be assigned a currency if another slot in the CEP card has already been assigned to that currency.”	Y
CPN	“On indication received at the CNG of a resource allocation expiry the CNG shall delete all residual data associated with the invocation of the resource.”	
	“It shall be possible to configure the CNG (e.g. firmware downloading) according to the subscribed services. This operation may be performed when the CNG is connected to the network for the first time, for each new service subscription/modification, or for any technical management (e.g. security, patches, etc.).”	Y
GPS	“The back-end systems (multiple back-end systems may exist for a single card), which communicate with the cards, perform the verifications, and manage the off-card key databases, also shall be trusted.”	
	“If an Application implicitly selectable on specific logical channel(s) of specific card I/O interface(s) is deleted, the Issuer Security Domain becomes the implicitly selectable Application on that logical channel(s) of that card I/O interface(s).”	Y

Practice Exercise 3
Draw Histogram and cluster

Draw Histogram

Table 2.5: NFR dataset, broken down by project and requirements type [14].

		Project ID															
Requirement Type	*Label*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	*Total*
Availability	A	1	1	1	0	2	1	0	5	1	1	1	1	1	1	1	**18**
Legal	L	0	0	0	3	3	0	1	3	0	0	0	0	0	0	0	**10**
Look-and-Feel	LF	1	2	0	1	3	2	0	6	0	7	2	2	4	3	2	**35**
Maintainability	MN	0	0	0	0	0	3	0	2	1	0	1	3	2	2	2	**16**
Operational	O	0	0	6	6	10	15	3	9	2	0	0	2	2	3	3	**61**
Performance	PE	2	3	1	2	4	1	2	17	4	4	1	5	0	1	1	**48**
Scalability	SC	0	1	3	0	3	4	0	4	0	0	0	1	2	0	0	**18**
Security	SE	1	3	6	6	7	5	2	15	0	1	3	3	2	2	2	**58**
Usability	US	3	5	4	4	5	13	0	10	0	2	2	3	6	4	1	**62**
Total NFRs		8	15	21	21	37	44	8	71	8	15	10	20	19	16	12	*326*
Functional	F	20	11	47	25	36	26	15	20	16	38	22	13	3	51	15	**358**
Total		28	26	68	47	73	70	23	91	24	53	32	33	22	67	127	*684*

Practice Exercise 4

Highlight/Circle best Keyword/Write one-word on your own

Given Statistics and Description

The *Quality Attributes* (NFR) dataset [4], also known as the *PROMISE* corpus, is a compilation of requirements specifications for 15 software projects developed by MS students at DePaul University as a term project for a Requirements Engineering course [14]. The dataset consists of 326 non-functional requirements (NFRs) of nine types and 358 functional requirements (FRs). Table 2.5 tabulates the distribution of requirement types among the 15 projects, and Table 2.6 provides examples of each type of requirement.

UNIVERSITY OF KARACHI

4

Highlight/Identify the type of requirement. Circle best key words which help you in identifying requirement type

Table 2.6: Examples of requirements of different types from NFR.

Label	*Requirements Text*
A	"The RFS system should be available 24/7 especially during the budgeting period. The RFS system shall be available 90% of the time all year and 98% during the budgeting period. 2% of the time the system will become available within 1 hour of the time that the situation is reported."
L	"The System shall meet all applicable accounting standards. The final version of the System must successfully pass independent audit performed by a certified auditor."
LF	"The website shall be attractive to all audiences. The website shall appear to be fun and the colors should be bright and vibrant."
MN	"Application updates shall occur between 3AM and 6 AM CST on Wednesday morning during the middle of the NFL season."
O	"The product must work with most database management systems (DBMS) on the market whether the DBMS is colocated with the product on the same machine or is located on a different machine on the computer network."

OF
HI

5

Highlight/Identify the type of requirement. Circle best key words which help you in identifying requirement type

PE	"The search for the preferred repair facility shall take no longer than 8 seconds. The preferred repair facility is returned within 8 seconds."
SC	"The system shall be expected to manage the nursing program curriculum and class/ clinical scheduling for a minimum of 5 years."
SE	"The product shall ensure that it can only be accessed by authorized users. The product will be able to distinguish between authorized and unauthorized users in all access attempts."
US	"If projected the data must be readable. On a 10x10 projection screen 90% of viewers must be able to read Event / Activity data from a viewing distance of 30."
F	"System shall automatically update the main page of the website every Friday and show the 4 latest movies that have been added to the website."

UNIVERSITY OF KARACHI

Complete missing values in table and show all working steps

Practice Exercise 5
Naïve Bayes

Step 1: Convert the data set into a frequency table using Predictors and Response variable

Variable : Weather   States{ sunny, overcast, rainy…….}

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

More Variables and More States

Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

Frequency Table for Predictors

Frequency Table		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1

		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1

		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3

Practice Exercise 6
Naïve Bayes

Step 2: Create Likelihood table by finding the probabilities

4

Calculate Likelihood probabilities

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Likelihood table				
Weather	No	Yes		
Overcast		4	=4/14	0.29
Rainy	3	2	=5/14	0.36
Sunny	2	3	=5/14	0.36
All	5	9		
	=5/14	=9/14		
	0.36	0.64		

Likelihood probability for Outlook

$$P(x \mid c) = P(Sunny \mid Yes) = 3/9 = 0.33$$

Frequency Table		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

Likelihood Table		Play Golf		
		Yes	No	
Outlook	Sunny	3/9	2/5	5/14
	Overcast	4/9	0/5	4/14
	Rainy	2/9	3/5	5/14
		9/14	5/14	

$$P(x) = P(Sunny) = 5/14 = 0.36$$

$$P(c) = P(Yes) = 9/14 = 0.64$$

UNIVERSITY OF KARACHI

Likelihood probability for other attributes

		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1

		Play Golf	
		Yes	No
Humidity	High	3/9	4/5
	Normal	6/9	1/5

		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1

		Play Golf	
		Yes	No
Temp.	Hot	2/9	2/5
	Mild	4/9	2/5
	Cool	3/9	1/5

		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3

		Play Golf	
		Yes	No
Windy	False	6/9	2/5
	True	3/9	3/5

Practice Exercise 7
Naïve Bayes

Step 3: Calculate Posterior Probability of each class

Posterior Probability of Playing Golf (yes)

$$P(x \mid c) = P(Sunny \mid Yes) = 3/9 = 0.33$$

Frequency Table		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

Likelihood Table		Play Golf		
		Yes	No	
Outlook	Sunny	3/9	2/5	5/14
	Overcast	4/9	0/5	4/14
	Rainy	2/9	3/5	5/14
		9/14	5/14	

$$P(x) = P(Sunny)$$
$$= 5/14 = 0.36$$

$$P(c) = P(Yes) = 9/14 = 0.64$$

Posterior Probability: $P(c \mid x) = P(Yes \mid Sunny) = 0.33 \times 0.64 \div 0.36 = 0.60$

UNIVERSITY OF KARACHI

Posterior Probability of Not Playing Golf (No)

Frequency Table		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

$$P(x \mid c) = P(Sunny \mid No) = 2/5 = 0.4$$

		Play Golf		
		Yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
		9	5	14

$$P(x) = P(Sunny) = 5/14 = 0.36$$

$$P(c) = P(No) = 5/14 = 0.36$$

Posterior Probability: $P(c \mid x) = P(No \mid Sunny) = 0.40 \times 0.36 \div 0.36 = 0.40$

How they calculate final posterior probability?

Redo the same problem, do reasoning on final posterior probability and submit

[Naive Bayesian (saedsayad.com)](saedsayad.com)

Bayes Theorem

There are three kinds of lies: lies, damned lies and statistics.
—Mark Twain

The word "Bayesian" traces its origin to the 18th century and English Reverend Thomas Bayes, who along with Pierre-Simon Laplace was among the first thinkers to consider the laws of chance and randomness in a quantitative, scientific way. Both Bayes and Laplace were aware of a relation that is now known as Bayes Theorem:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} \propto p(x|\theta)p(\theta). \tag{1.1}$$

The proportionality $\propto$ in Eq. (1.1) signifies that the $1/p(x)$ factor is constant and may be ignored when viewing $p(\theta|x)$ as a function of θ. We can decompose Bayes' Theorem into three principal terms:

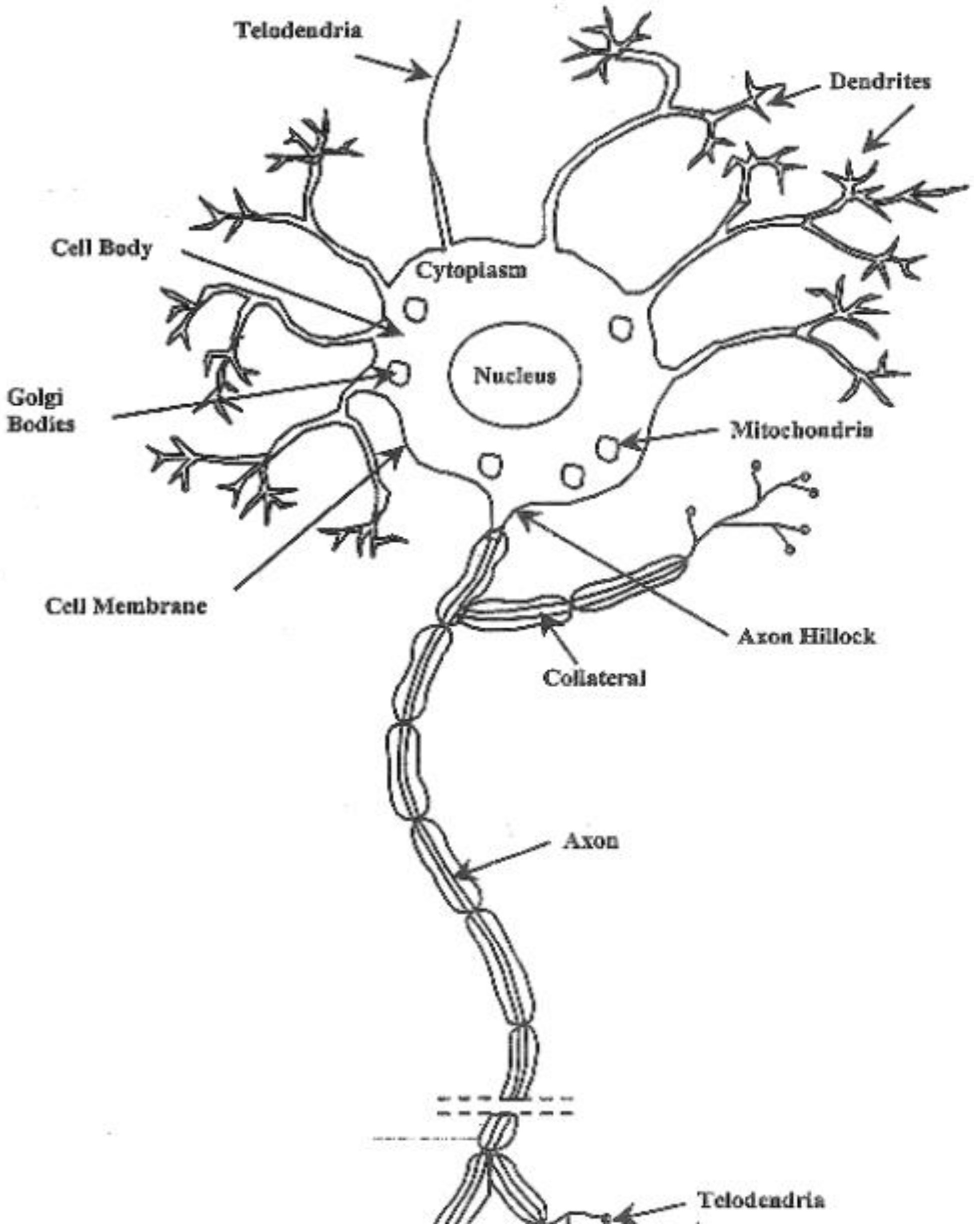
$p(\theta\|x)$	posterior
$p(x\|\theta)$	likelihood
$p(\theta)$	prior

Intro to NN and Perceptron

Practice Exercise 8

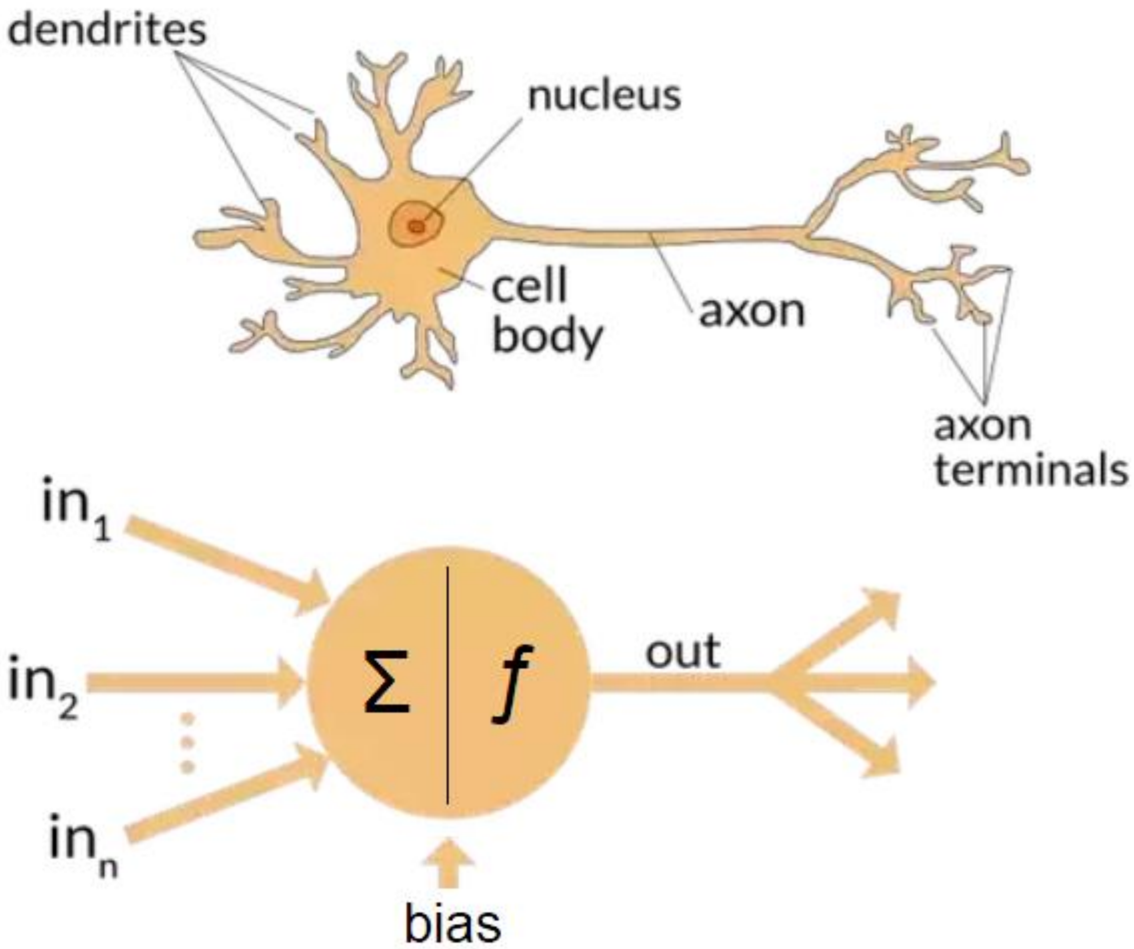
Identify Linear vs. Non-Linear

Structure of Biological Neuron



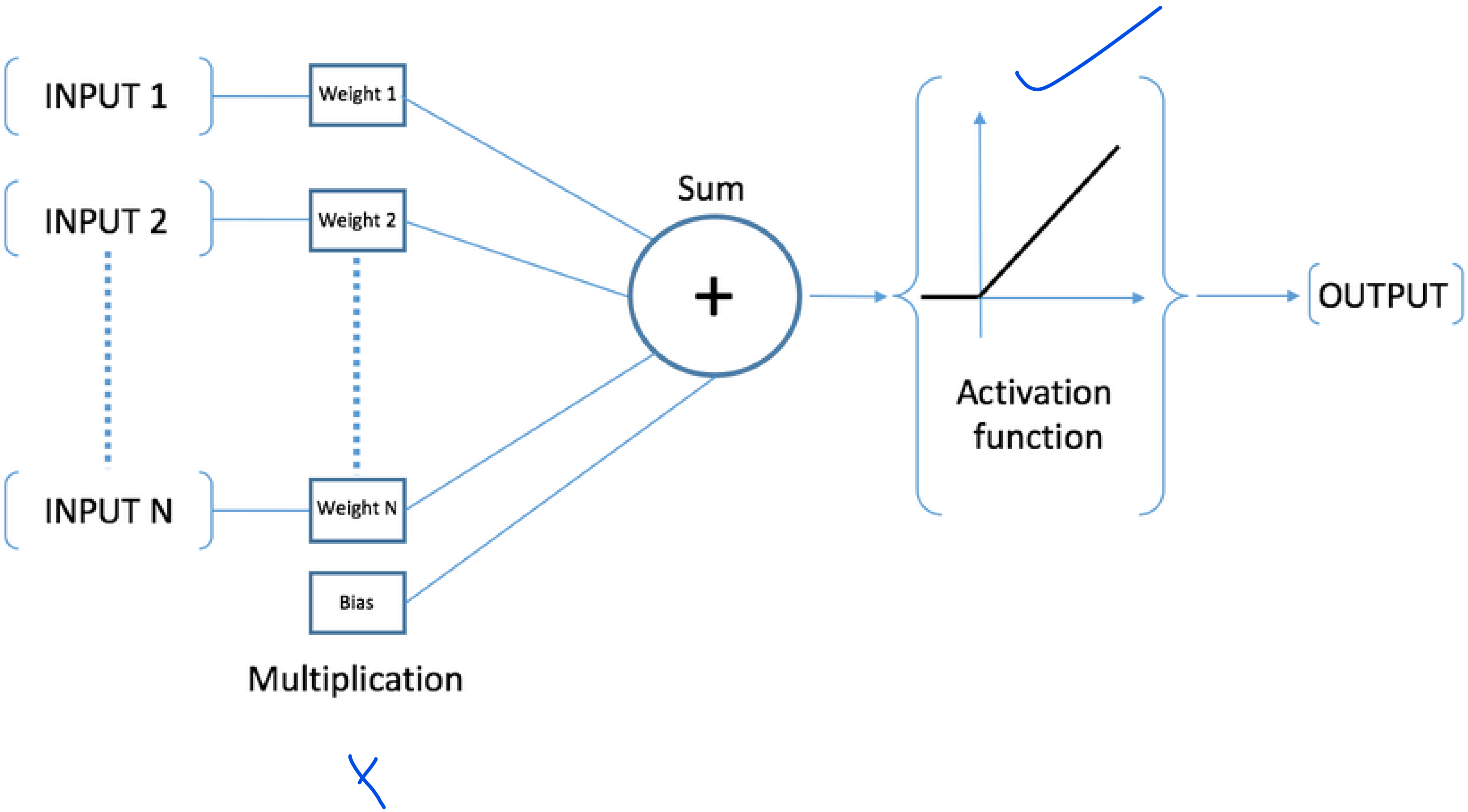
UNIVERSITY OF KARACHI

Biological vs. Artificial Neuron



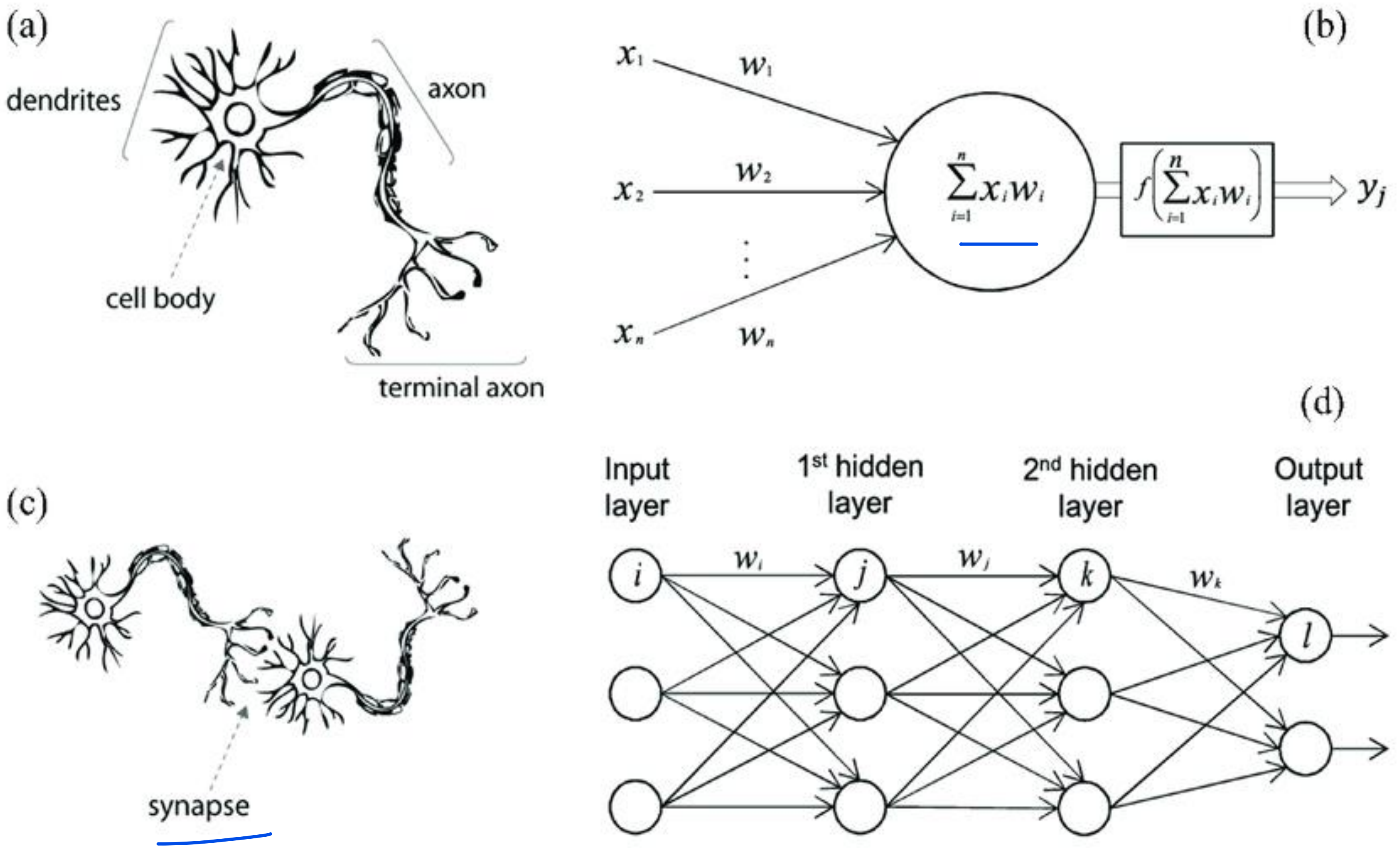
UNIVERSITY OF KARACHI

Sum vs. Non-linear Activation



UNIVERSITY OF KARACHI

Sum: Dot Product -> Sum of Products



Classification: Linearly vs. Non-Linearly separable

XOR Truth Table

Input 1	Input 2	Output
0	0	0
0	1	1
1	0	1
1	1	0

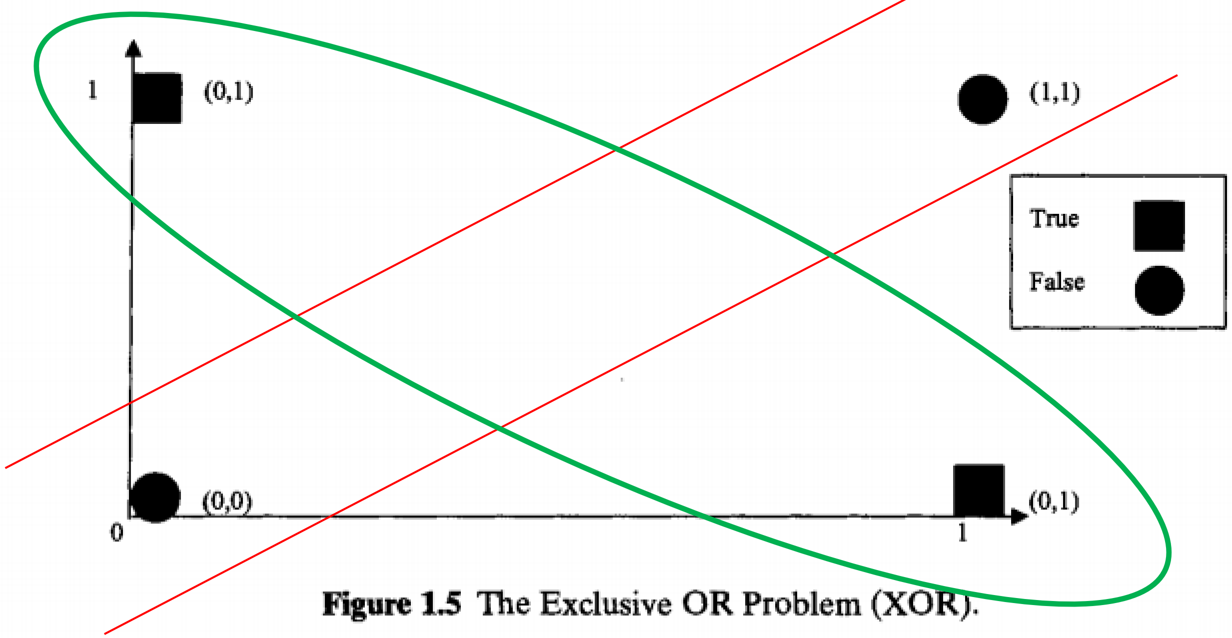
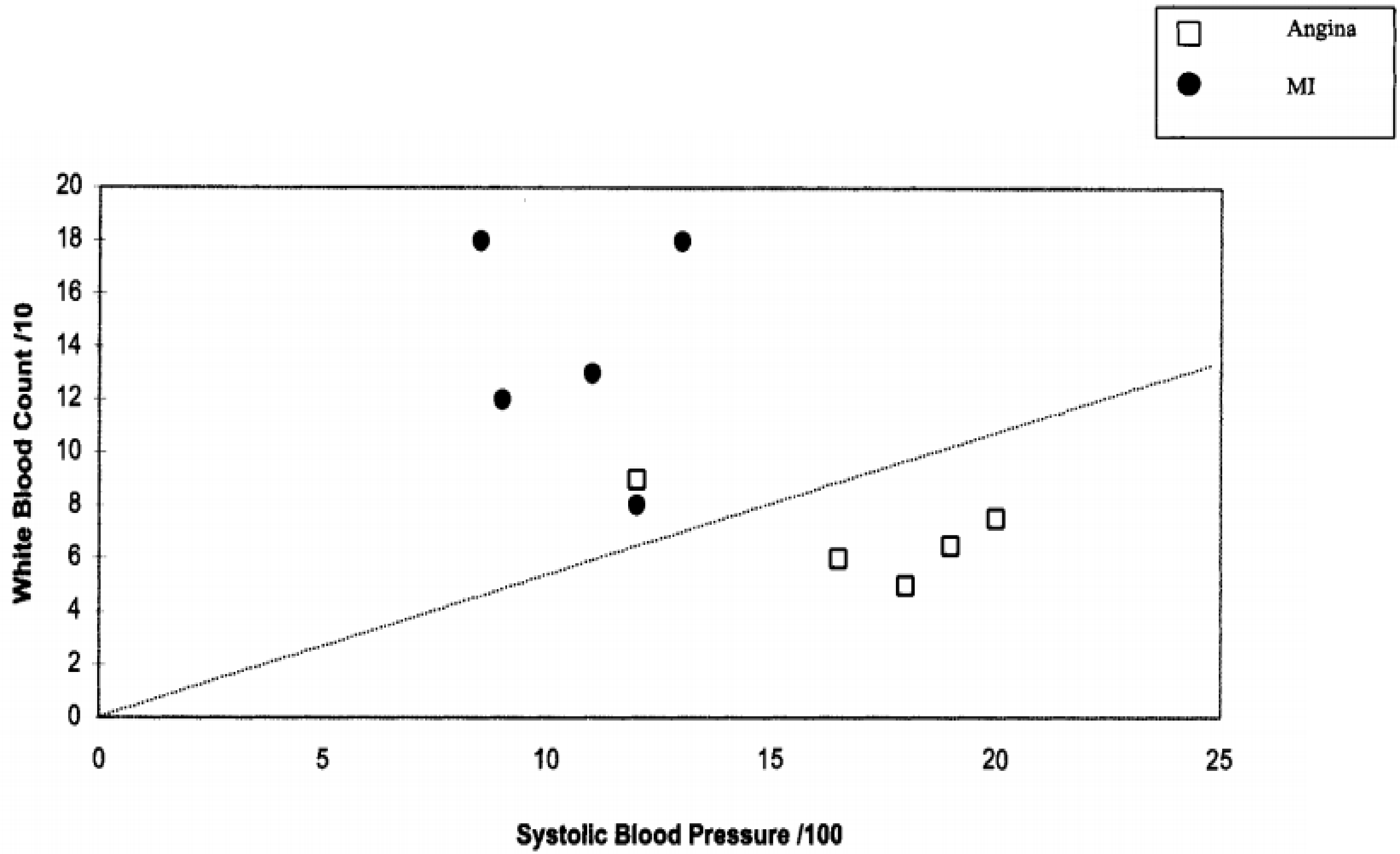


Figure 1.5 The Exclusive OR Problem (XOR).

UNIVERSITY OF KARACHI

2-Class Problem: Linear or Non-Linear



UNIVERSITY OF KARACHI

Intro to NN and Perceptron

Practice Exercise 9

Calculate dot product and take decision to update weight

Perceptron Learning Approach

A general algorithm for supervised learning follows:

Make an initial guess for each component of **w.**

Select a training set of data.

For each vector in the training set:

Compute D(**x**)

If D(**x**) > *0 and* **x** ε *class 1 or* D(**x**) < *0 and* **x** ε *class 2, do not adjust* **w**

If D(**x**) > *0 and* **x** ε *class 2 adjust* **w** *according to rule 1*

If D(**x**) < *0 and* **x** ε *class 1 adjust* **w** *according to rule 2*

Until **w** *does not change (or until criterion function is minimized).*

UNIVERSITY OF KARACHI

2-Class Problem: Linear or Non-Linear

TABLE 1.1 Feature Vector Values for Differentiation between Myocardial Infarction (MI) and Angina

Feature Vector	Diagnosis	Systolic Blood Pressure	White Blood Count
$\mathbf{x}_1$	MI	110	13,000
$\mathbf{x}_2$	MI	90	12,000
$\mathbf{x}_3$	MI	85	18,000
$\mathbf{x}_4$	MI	120	8,000
$\mathbf{x}_5$	MI	130	18,000
$\mathbf{x}_6$	Angina	180	5,000
$\mathbf{x}_7$	Angina	200	7,500
$\mathbf{x}_8$	Angina	165	6,000
$\mathbf{x}_9$	Angina	190	6,500
$\mathbf{x}_{10}$	Angina	120	9,000

UNIVERSITY OF KARACHI

Initial Guess Weight Vector and dot-product

$$D(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} = w_1 x_2 + w_2 x_2$$

$\mathbf{t}_1 = (11.0, 13.0)$ (vector $\mathbf{x}_1$, class 1)
$\mathbf{t}_2 = (18.0, 5.0)$ (vector $\mathbf{x}_6$, class 2)

$\mathbf{t}_3 = (9.0, 12.0)$ (vector $\mathbf{x}_2$, class 1)
$\mathbf{t}_4 = (20.0, 7.5)$ (vector $\mathbf{x}_7$, class 2)

We will make an initial guess for each weight as $w_1 = -0.3, w_2 = 1.0$. Initially, we substitute vector $\mathbf{t}_1$ into Eq. (1.5):

$$D(\mathbf{t}_1) = -0.3\ (11.0)\ +1.0(13) > 0; \text{therefore } y(t) = 1$$

$\mathbf{t}_1$ belongs to class 1; therefore $d(t) = 1$

Continue till convergence

$$D(\mathbf{t}_4) = -0.3(20.0) + 1.0(7.5) > 0, y(t) = 1$$

$\mathbf{t}_4$ belongs to class 2

Therefore, substituting into Eq. (1.6)

$$w_1(1) = -0.3 + 0.01[(-1 - (1)] \, 20.0 = -0.7$$
$$w_2(1) = 1.0 + 0.01[-1 - (1)]7.5 = 0.85$$

The process must then begin again with $\mathbf{t}_1$ and continue until all vectors are classified correctly. After completion of this process, the resulting weights are:

$$w_1 = -0.7$$
$$w_2 = 0.85$$

Our decision surface is

$$D(\mathbf{x}) = -0.7x_1 + 0.85x_2 \tag{1.7}$$

UNIVERSITY OF KARACHI

Lecture #06
NN

$\bullet \in \omega_1$
$\circ \in \omega_2$

2017 Week 5

February

2 THURSDAY 3 FRIDAY 4 SATURDAY

A.M. Gonzales 2nd Edition Example Object Recognition

Training Sets:

$\omega_1 = \{ (0,0,1)^T, (0,1,1)^T \}$ $\eta = 1$

$\omega_2 = \{ (1,0,1)^T, (1,1,1)^T \}$

Build Training Table just like we did in Good Tiles and Bad tiles Example to demonstrate Convergence.

LUNCH 5 SUNDAY

P.M. Let $\bar{W}(1) = 0$

Input $\bar{u}_k$	W_k^T	OK(?) ω_1/ω_2	$W_k^T \bar{u}_k$	Apply Perceptron Rule	$\bar{w}_{k+1}$
(0,0,1)	(0,0,0)	ω_1	0	$W_{k+1} = W_k$	0,0,0
(0,1,1)	(0,0,0)	ω_1	0	$W_{k+1} = W_k$	0,0,0
(1,0,1)	(0,0,0)	ω_2	0	$W_{k+1} = W_k - \eta u_k$	−1,0,−1
(1,1,1)	(−1,0,−1)	ω_2	−2	$W_{k+1} = W_k$	(−1,0,−1)

EVENING

2nd Pass

Input $\bar{u}_k$	W_k^T	OK(?) ω_1/ω_2	$W_k^T \bar{u}_k$	Apply Perceptron Rule	$\bar{w}_{k+1}$
(0,0,1)	(−1,0,−1)	ω_1	−1	$W_{k+1} = W_k + \eta u_k$	(−1,0,0)
(0,1,1)	(−1,0,0)	ω_1	0	$W_{k+1} = W_k$	(−1,0,0)
(1,0,1)	(−1,0,0)	ω_2	−1	$w_{k+1} = w_k$	(−1,0,0)
(1,1,1)	(−1,0,0)	ω_2	−1	$w_{k+1} = w_k$	(−1,0,0)

UNIVERSITY OF KARACHI

February

Week 7 2017

13 MONDAY 14 TUESDAY 15 WEDNESDAY

Weight update Rule / Learning

- Perceptron
  - Error is constant
- Gradient descent / Δ Rule
  - Error is function / LMS / surface
- BP (MLP)

$$\Delta w_i = \eta (t-o) x_i$$

$$\Delta w_i = \eta \bar{x}_i$$

OR

$$\Delta W_K = \eta \bar{u}_K$$

$t - o = 0$ (Δw_i is zero)

$t - o = +2$ when $t = +1$, $o = -1$ (weights ↑se)

$t - o = -2$ when $t = -1$, $o = 1$ (weights ↓se)

$$\boxed{\text{Error} = t - O}$$

$$\text{Error} = \frac{1}{2} \sum_{d \in D} (t_d - o_d)^2 = \frac{1}{2} \sum e^2$$

D : set of training examples.

$$\Delta w_i = -\eta \frac{\partial E}{\partial w_i}$$

$$\boxed{\Delta w_i = -\eta \left[\sum_{d \in D} (t_d - o_d) x_{id} \right]}$$

$$E(\bar{w}) = \frac{1}{2} \sum_{d \in D} \sum_{k \in \text{outputs}}$$

BP

$$\Delta w_{ji} = -\eta \frac{\partial E_d}{\partial w_{ji}}$$

$$\boxed{\Delta w_{ji} = -\eta (t_j - o_j) o_j (1 - o_j) x_{ji}}$$

using chain rule on $\frac{\partial E_d}{\partial w_{ji}}$

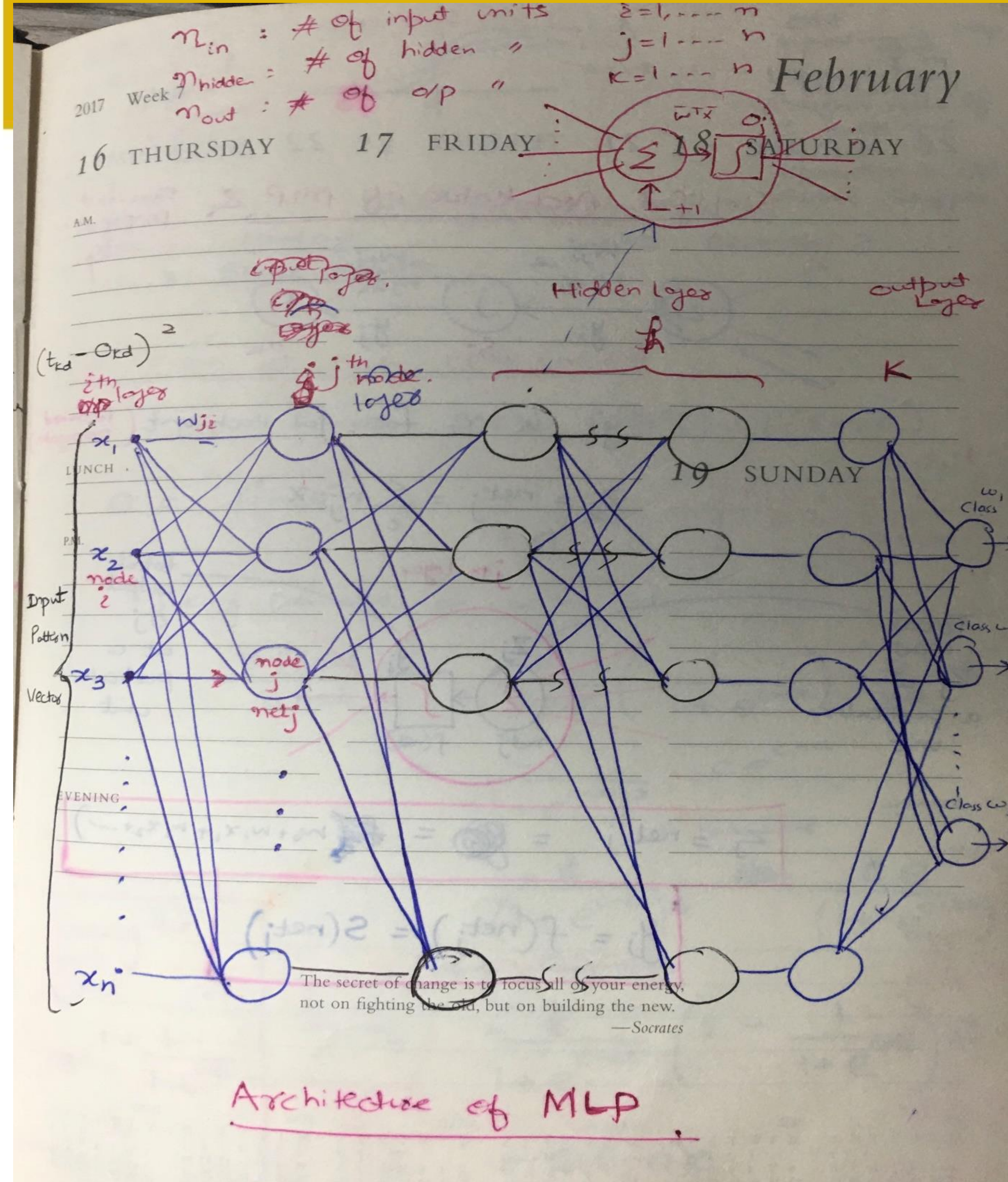
$$\boxed{\Delta w_{ji} = +\eta \delta_j x_{ji}}$$

NOTES

MLP consist of same i/p, hidden and o/p layers. But it employ BP for training

JAN. FEB. MARCH APRIL MAY JUNE

UNIVERSITY OF KARACHI



UNIVERSITY OF KARACHI

Declaration

✓ Submit all practice question in this presentation

☐ I don't understand Bayes and NN Intro

☐ If not, What is the main hindrance to start work?

☐ If Yes, share your experience/practice/work