

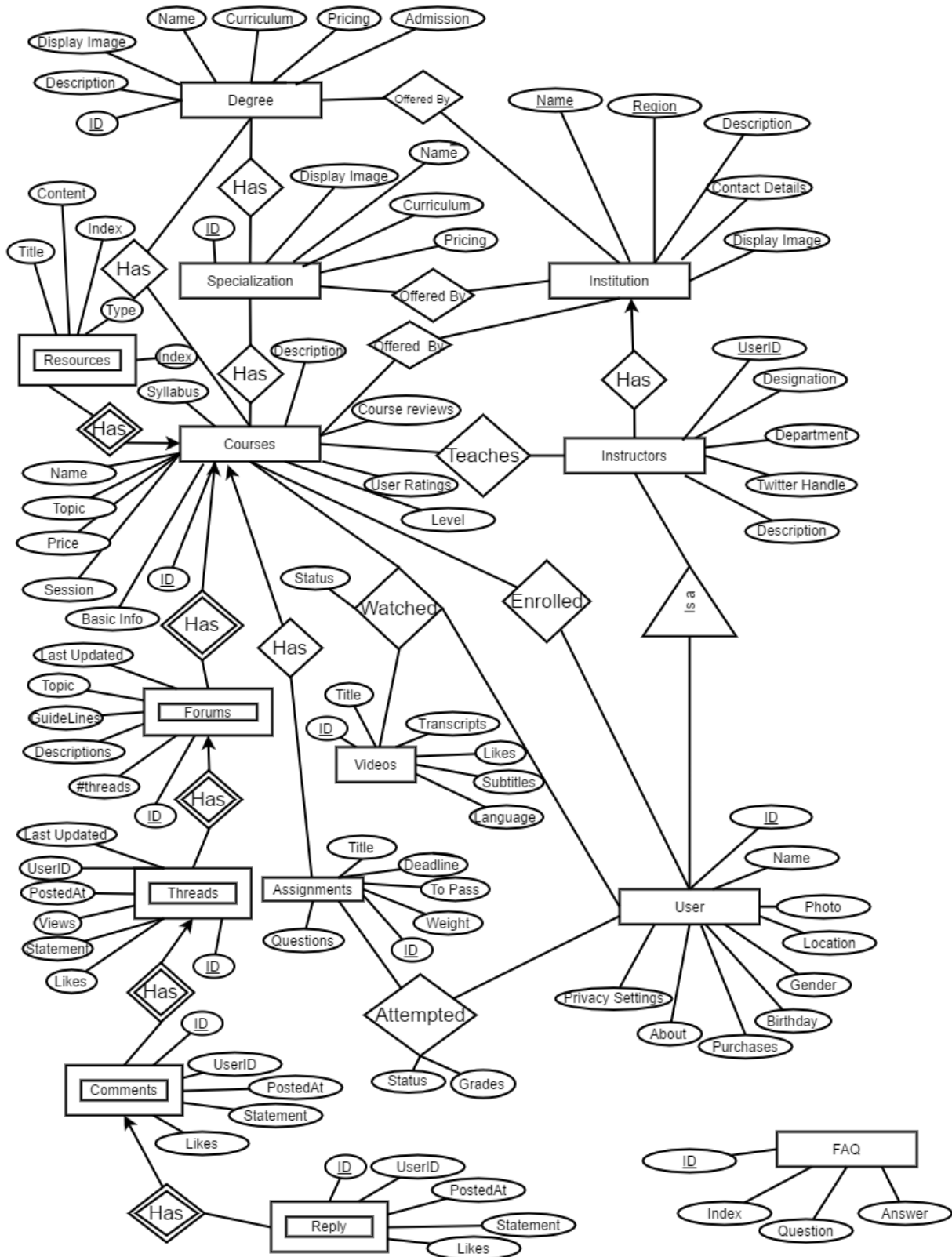
# Assignment 1

Harshdeep Gupta, 2013MT60597

Deepali Gupta, 2013MT60079

Due date: January 25, 2017, 11:55pm IST

### 1.1 ER Diagram



### ER Diagram

Table 1: Entities and Attributes

Entity	Attributes	Keys
Degree	ID, Name, Curriculum, Pricing, Admission, AboutTheProgram DisplayImage	ID
Specialization	ID, Name, Price, Description, DisplayImage	ID
Course	ID, Name, Topic, Price, Session, BasicInfo, Level, Language, DisplayImage, HowToPass, UserRatings, CourseReviews, AboutTheCourse, WhoIsThisCourseFor, Syllabus, CertificateOffering, Description	ID
Institution	Name, Region, Description, DisplayImage, Contact Details	Name, Region
User	ID, Name, Photo, Location, Gender, Birthday, AboutUser, Websites, Privacy, Purchases	ID
Instructor	UserID, Designation, Twitter Handle, Department, Description	UserID
Assignment	Index, Title, Deadline, ToPass, Weight, Grade, Questions	CourseID, Index
Video	ID, Title, Transcript, Likes, Subtitles, Resources, Language	ID
Forum	ID, Topic, Guidelines, Description, No of threads, Last updated	ID
Thread	ID, UserID, Title, Post, Views, Likes, Last Posted, Content	ID
Comment	ID, UserID, Statement, Likes, PostedAt	ID
Reply	ID, UserID, Statement, Likes, PostedAt	ID
FAQ	ID, Index, Question, Answer	ID
Resource	Index, Type, Title, Content, Last Updated	CourseID, Index

## 1.2 Set of relations

- Degree(ID, Name, Curriculum, Pricing, Admission, AboutTheProgram, DisplayImage)
- Specialization(ID, Name, Price, Description, DisplayImage)
- Course(ID, Name, Topic, Price, Session, BasicInfo, Level, Language, DisplayImage, HowToPass, UserRatings, CourseReviews, AboutTheCourse, WhoIsThisCourseFor, Syllabus, CertificateOffering, Description)
- Institution(Name, Region, Description, DisplayImage, Contact Details)
- User(ID, Name, Photo, Location, Gender, Birthday, AboutUser, Websites, Privacy, Purchases)
- Instructor(UserID, Designation, Twitter Handle, Department, Description)
- Assignment(CourseID, Index, Title, Deadline, ToPass, Weight, NoOfQuestions)
- Video(ID, CourseID, Title, Transcript, Likes, Subtitles, Resources, Language)
- Forum(ID, CourseID, Topic, Guidelines, Description, No of threads, Last updated)
- Thread(ID, ForumID, UserID, Title, Post, Views, Likes, Last Posted, Content)
- Comment(ID, ThreadID, UserID, Statement, Likes, PostedAt)
- Reply(ID, CommentID, UserID, Statement, Likes, PostedAt)
- FAQ(ID, Index, Question, Answer)
- Resource(CourseID, Index, Type, Title, Content, Last Updated)
- DegreeSpecs(DegreeID, SpecID)
- SpecCourses(SpecID, CourseID)
- DegreeCourses(DegreeID, CourseID)
- InstFaculty(InstName, InstRegion, InstructorID)

- InstDegree(InstName, InstRegion, DegreeID)
- InstSpec(InstName, InstRegion, SpecID)
- InstCourse(InstName, InstRegion, CourseID)
- EnrolledIn(CourseID, UserID)
- Teaches(CourseID, InstructorID)
- Watched(VideoID, UserID, Status)
- Attempted(CourseID, AssignIndex, UserID, Status, Grade)

### 1.3 Keys and FDs

All the keys are underlined in the previous section. Below are the FD's

- Institution: Contact Details  $\rightarrow$  Name , Region
- Forum: (CourseID, Topic)  $\rightarrow$  ForumID
- Thread: (ForumID, UserID, Title)  $\rightarrow$  ThreadID

### 1.4 Sample Data

Table 2: Course

ID	Name	Level	Price	Session	HowToPass	UserRatings
1	Neural Networks	NULL	\$95	Jan23-April23	Pass all assignments	4.5
2	Game Theory	Beginner	\$50	Feb18-May20	Pass all assignments	4.6
3	Advanced Writing	NULL	\$40	Jan23-April23	Pass all assignments	4.6
4	Algebra	Beginner	\$80	Jan23-Mar07	Pass all assignments	4.5
5	Hadoop Platform	Intermediate	\$60	Jan23-Feb30	Pass all assignments	NULL

Table 3: User

ID	Name	Location	Gender	Birthday	Privacy
1	Deepali Gupta	New Delhi, India	Female	23/05/1995	OnlyMe
2	Vladimir Diaz	New York, USA	Male	02/01/1984	The Coursera Community
3	Santiago Torres Arias	Mexico City, Mexico	Male	06/07/1989	Everyone on the Web
4	Ghada Al-mashaqbeh	Chicago, USA	Female	25/08/1979	OnlyMe

Neural Networks for Machine Learning - University of Toronto | Coursera - Google Chrome

Assignment 1 - COL362 Ass1 - Lecture17.pdf - Inbox - deepalg - WhatsApp - Neural Network - Bold, italics and - Deepall

Secure | <https://www.coursera.org/learn/neural-networks#>

# Neural Networks for Machine Learning

Overview

Syllabus

FAQs

Creators

Ratings and Reviews

Neural Networks for Machine Learning


**Enroll Now**  
Started Jan 23


Financial Aid is available for learners who cannot afford the fee. [Learn more and apply.](#)




**About this course:** Learn about artificial neural networks and how they're being used for machine learning, as applied to speech and object recognition, image segmentation, modeling language and human motion, etc. We'll emphasize both the basic algorithms and the practical tricks needed to get them to work well.

[More](#)

**Created by:** University of Toronto



 **Taught by:** Geoffrey Hinton, Professor  
Department of Computer Science

 <b>Language</b>	English
 <b>How To Pass</b>	Pass all graded assignments to complete the course.
 <b>User Ratings</b>	★★★★★ Average User Rating 4.5 <a href="#">See what learners said</a>

Lect1.pdf ^ Lect2.pdf ^ Lect3.pdf ^ Lect4.pdf ^ Lect5.pdf ^ **Show all** x

Course Introduction Page

Neural Networks for Machine Learning - University of Toronto | Coursera - Google Chrome

Assignment 1 - COL362 Ass1 - Lecture17.pdf - Inbox - deepalg - WhatsApp - Neural Network - Bold, italics and - Deepall

Secure | <https://www.coursera.org/learn/neural-networks#>

# Neural Networks for Machine Learning

Overview

Syllabus

FAQs

Creators

Ratings and Reviews

Neural Networks for Machine Learning

**Enroll Now**  
Started Jan 23

Financial Aid is available for learners who cannot afford the fee. [Learn more and apply.](#)

**WEEK 1**

**Introduction**  
Introduction to the course - machine learning and neural nets  
5 videos, 8 readings [expand](#)

[View Full Syllabus](#)

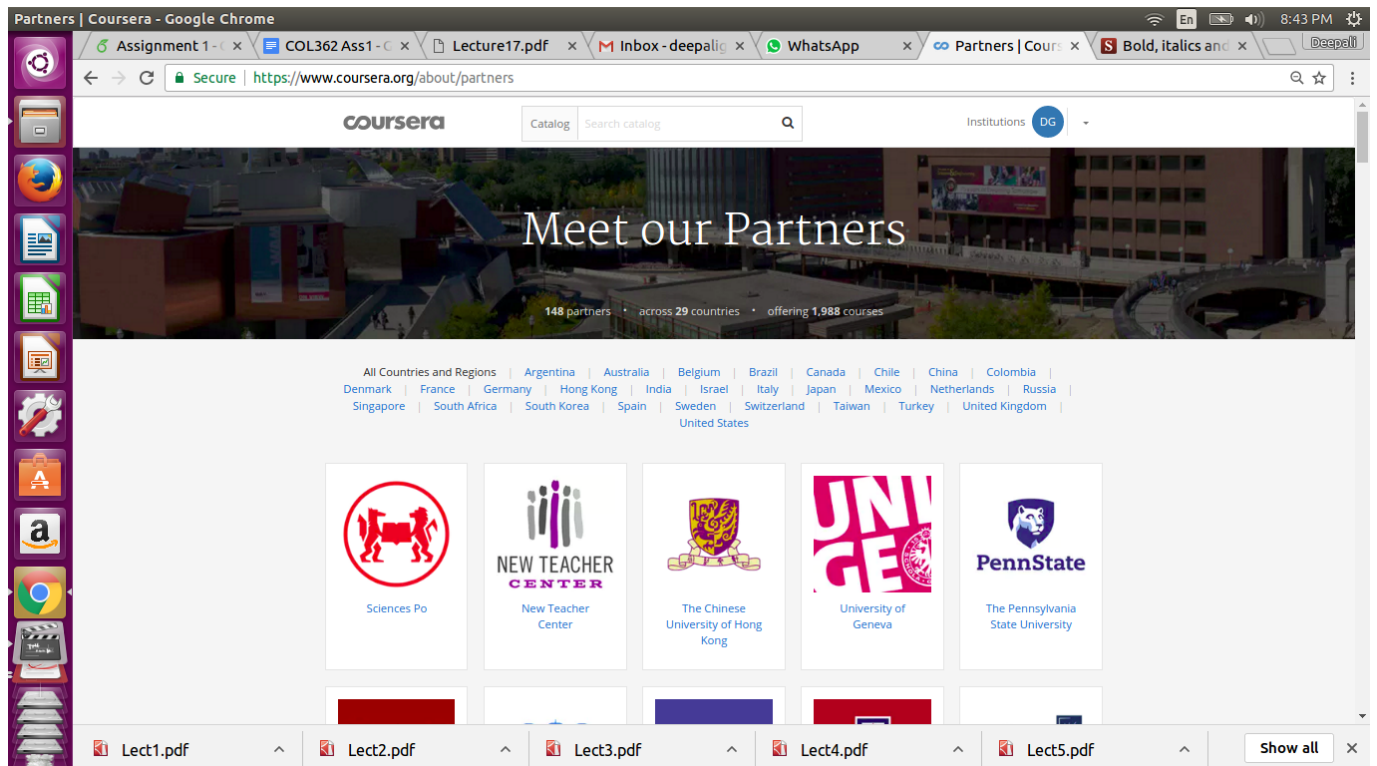
**Enroll Now**

**FAQs**

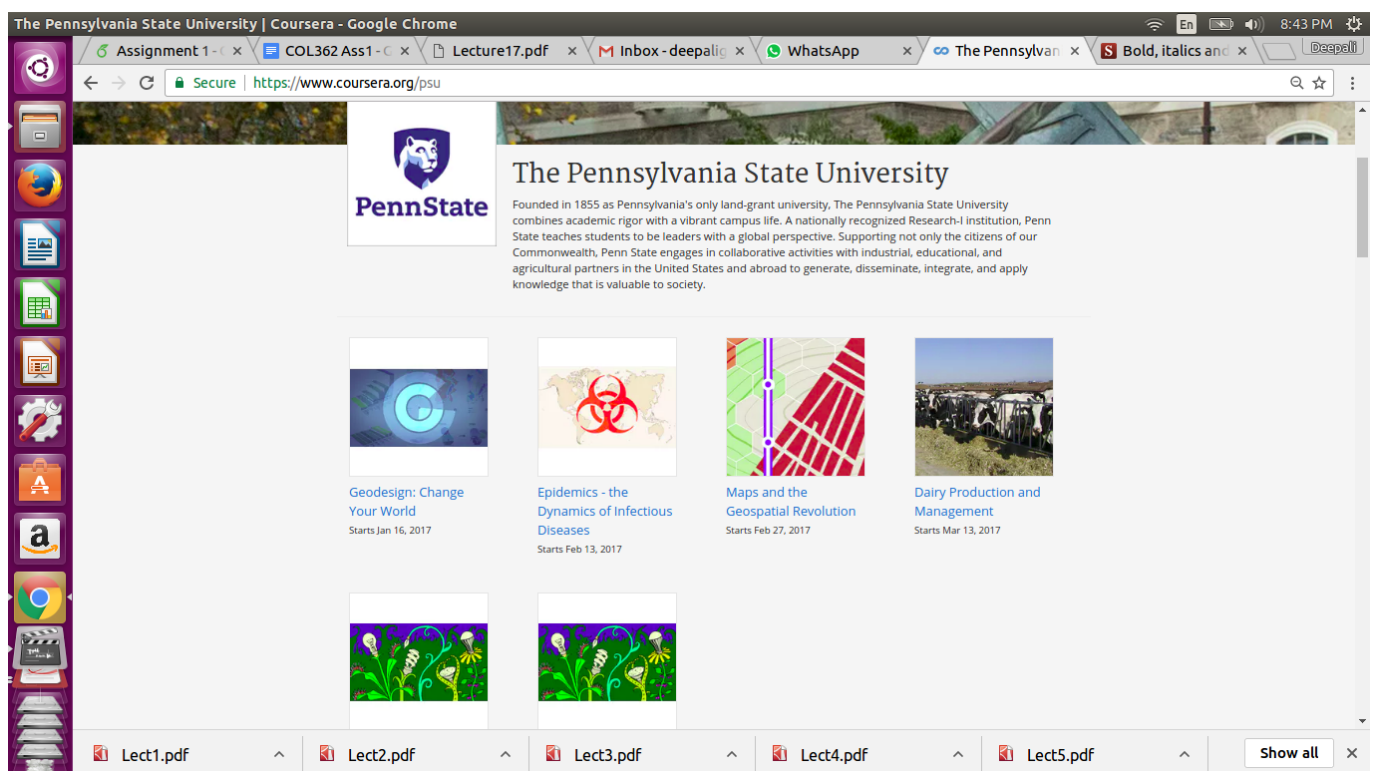
- When will I have access to the lectures and assignments?
- What if I need additional time to complete the course?
- What will I get if I pay for this course?
- Can I take this course for free?
- What is the refund policy?
- Is financial aid available?

Lect1.pdf ^ Lect2.pdf ^ Lect3.pdf ^ Lect4.pdf ^ Lect5.pdf ^ **Show all** x

Course Introduction Page



Universities Page



University Details Page

Neural Networks for Machine Learning | Coursera - Google Chrome

Assignment 1 - Overview x | Inbox - deepaligupta x | WhatsApp x | Neural Networks for Machine Learning x

Secure | https://www.coursera.org/learn/neural-networks/home/welcome

**coursera** DG

**Course Home**

- Week 1
- Week 2
- Week 3
- Week 4
- Week 5
- Week 6
- Week 7
- Week 8
- Week 9
- Week 10
- Week 11
- Week 12
- Week 13

**Introduction**

Videos 43 min left

Readings 1h 20m left

**REQUIRED**

Quiz Lecture 1 Quiz 12 min

**GRADE**

**DUE** Jan 29

**WEEK 2** Estimated time: 1h 7m

**The Perceptron learning procedure**

Videos 41 min left

Readings 10 min left

**REQUIRED**

Quiz Lecture 2 Quiz 16 min

**GRADE**

**DUE** Feb 5

**WEEK 3** Estimated time: 1h 16m

**The backpropagation learning procedure**

Videos 42 min left

**REQUIRED**

Quiz Lecture 3 Quiz

**GRADE**

**DUE** Feb 12

Lect1.pdf Lect2.pdf Lect3.pdf Lect4.pdf Lect5.pdf

Show all

Help Center

Course Details Page

Lecture 1 Quiz | Coursera - Google Chrome

Assignment 1 - Overview x | Inbox - deepaligupta x | WhatsApp x | Lecture 1 Quiz | Coursera x

Secure | https://www.coursera.org/learn/neural-networks/exam/chgv2/lecture-1-quiz

**coursera** DG

Back to Week 1 Lessons Prev Next

**Lecture1**

- Syllabus and Course Logistics 10 min
- Lecture Slides (and resources) 10 min
- Why do we need machine learning? (13 min) 13 min
- What are neural networks? (8 min) 8 min
- Some simple models of neurons (8 min) 8 min
- A simple example of learning (6 min) 5 min
- Three types of learning (8 min) 7 min
- Quiz: Lecture 1 Quiz 6 questions

**Octave Installation**

**QUIZ**

**Lecture 1 Quiz**

6 Questions

**Deadline** Pass quiz by January 29, 11:59 PM PST.

**To Pass** Earn at least 80%. This quiz counts towards passing the course.

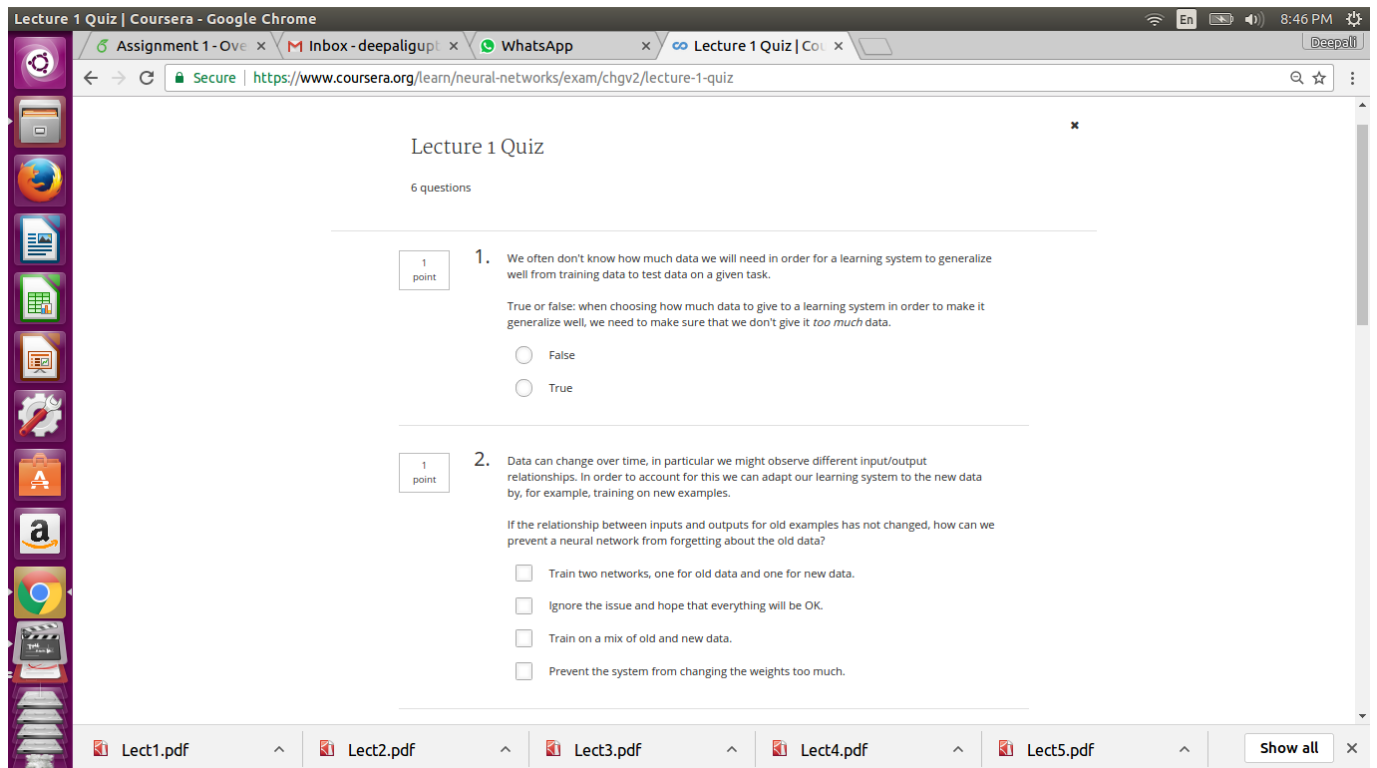
**Retakes** 3 attempts every 8 hours

Start Quiz

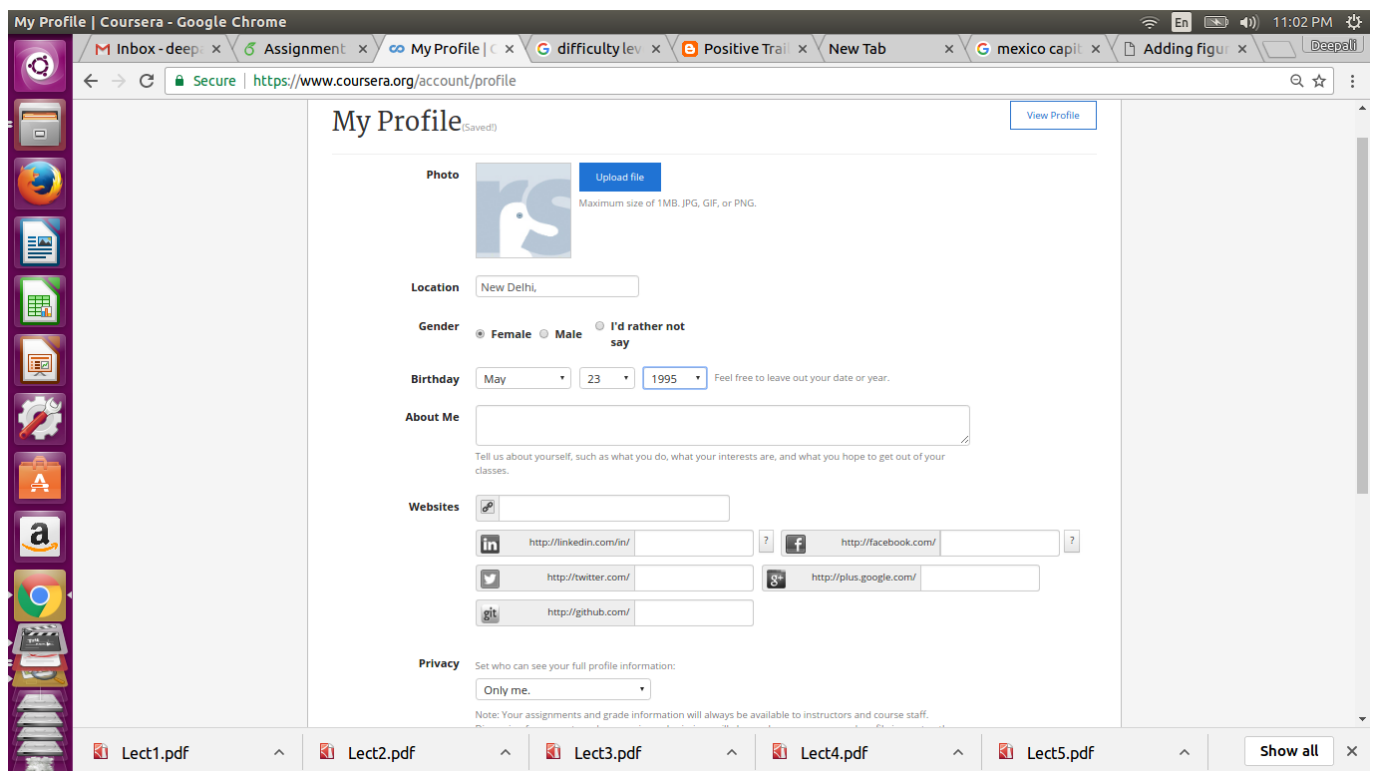
Lect1.pdf Lect2.pdf Lect3.pdf Lect4.pdf Lect5.pdf

Show all

Assignment Page

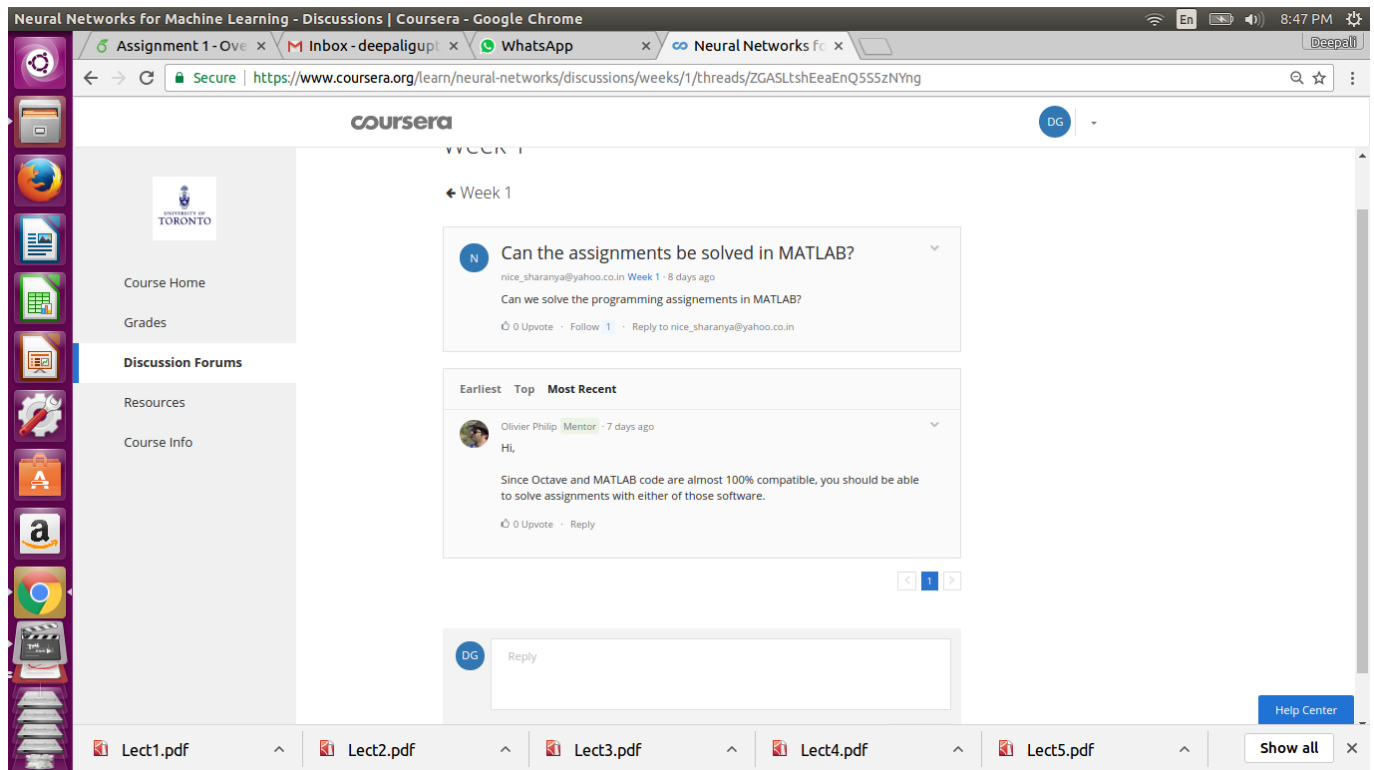


Assignment Questions Page



User Profile Page





Discussion Forum

## 1.5 List of various relationships

Table 4: Relationships

Type of relationship	Example
Weak Entity Set	(Assignment, Question), (Resource)
Non-Binary Relationship	(Course, User, Video)
Hierarchical Relationship	Instructor is a type of User
Type of relationship Constraint	All keys are unique and non null, User age must be in a valid range etc
Referential Constraint	ThreadID in comments refers to Thread, ForumID in thread refers to Forum, CourseID in thread refers to Course

## 2 Answer 2

Combining the attributes from tables: Course, User and Instructor, suppose we create a table with the following schema.

R(CourseID, Forum Topic, UserID, ThreadTitle, ThreadID, ForumID, ThreadViews, University Name, Contact Details)

For the purposes of example, we assume the following FD's in our data

- Contact Details  $\rightarrow$  Name
- (CourseID, Forum Topic)  $\rightarrow$  ForumID
- (ForumID, UserID, Thread Title)  $\rightarrow$  ThreadID
- (ThreadID)  $\rightarrow$  (Thread Title, Thread Views)

- $(\text{ForumID}) \rightarrow (\text{Forum Topic})$
- $(\text{ForumID}, \text{ThreadID}) \rightarrow (\text{UserID})$

We see that  $(\text{CourseID}, \text{ForumTopic}, \text{UserID}, \text{ThreadTitle}, \text{Contact Details})$  forms a key for this relation as its closure is the entire relation. Also, its minimal

**Decomposition into 2NF** In 2NF, no non prime attribute is dependent on proper subset of candidate key of the table. In the above relation, we see that the first three FD's violate this rule. So we decompose R into following relations :

- R1(Contact Details, UniName)
- R2(CourseID, Forum Topic, Forum ID)
- R3(ForumID, UserID, Thread Title, Thread Views, ThreadID)
- R4(CourseID, ForumTopic, UserID, Contact Details, Thread Title)

**Decomposition into 3 NF** In 3 NF, no non prime attribute can depend on another non prime attribute. We remove all the transitive dependencies.

- R2(a) (CourseID, ForumID)
- R2(b) (ForumID, ForumTopic)
- R3(a) (ThreadID, Thread Title, Thread Views)
- R3(b) (ForumID, UserID, ThreadID)

**Decomposition into BCNF** We see that the last FD is violating the BCNF criteria, so we modify relation 3(b).

- R3(b)(i) (ForumID, ThreadID, UserID)

## 3 Answer 3

### 3.1 Schema

Table 5: Dataset 1: Consumer Complaints

Field Name	Data Type
ComplaintID	bigint
Product	varchar(40)
Issue	varchar(80)
Company	varchar(20)
State	char(2)
TimelyResponse	varchar(3)

Table 6: Dataset 2: Most Recent Cohorts Scorecard Elements

Field Name	Data Type
UnitID	bigint
INSTNM	varchar(40)
City	varchar(20)
Insturl	varchar(40)
Npcurl	varchar(100)

Table 7: Dataset 3: US Chronic Disease Indicators

Field Name	Data Type
YearStart	char(4)
YearEnd	char(4)
LocationAbbr	char(2)
Topic	varchar(20)
Question	varchar(200)
Response	varchar(100)

## 3.2 Various insert modes

### 3.2.1 Bulk Load

Data was loaded from the csv files into the tables using the COPY command as follows:  
COPY consumer\_complaints FROM ' /Downloads/Consumer\_Complaints.csv' DELIMITER ',' csv;

### 3.2.2 Insert statements

A python script was written to load tuples from the csv file and place them into insert statements using psycopg2.

### 3.2.3 JDBC

A JDBC script was written to do a bulk insert.

## 3.3 Statistics

### Machine Configuration

- Processor : intel CORE i5
- RAM : 8 GB
- HDD : 1 TB
- SSD : 125 GB
- OS : Ubuntu 15.04
- JDBC : Java 8

Table 8: Data Insertion Statistics

Dataset No.	No. of tuples	Size(MB)	Bulk Load Time(ms)	Inserts Time(s)	JDBC Time(ms)
1	699224	59.5	3464.832	73	8649.273
2	53921	7	438.921	10	1357.538
3	237962	23	3837.433	35	7912.924

## 4 Answer 4

A python script was written to duplicate the data in the previous datasets by around 150 to 250 times.

Table 9: Data Insertion Statistics for Larger Datasets

Dataset No.	No. of tuples	Size(GB)	Bulk Load Time(min)	Inserts Time(min)	JDBC Time(min)
1	99047410	8.3	20	76	26
2	39439360	5.1	8	48	12
3	63535854	6.5	14	59	19