

# Assignment Project Exam Help

Machine learning lecture slides

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

## Overview

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

- ▶ Please use Piazza Live Q&A

# Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

- ▶ A “bird’s eye view” of machine learning
- ▶ About COMS 4771

# Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

# Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pro

Figure 1: Predict the bird species depicted in a given image.

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

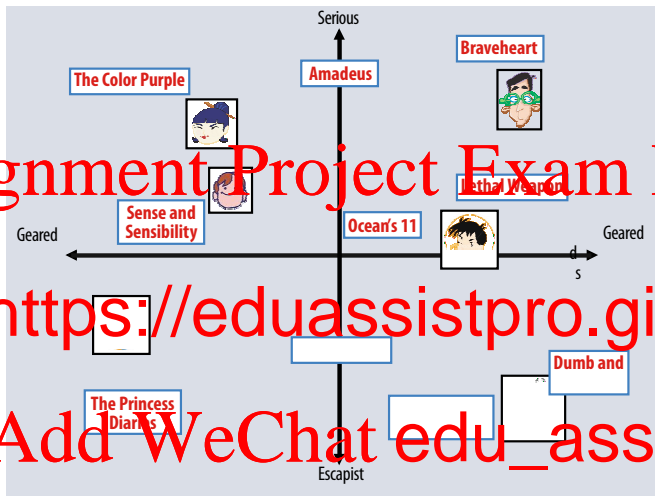
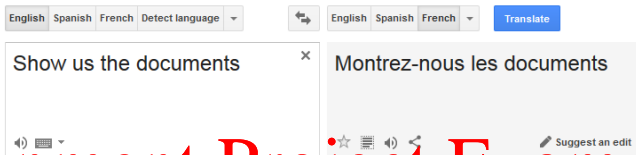


Figure 2: Predict how a given user would rate a given movie.



# Assignment Project Exam Help

Figure 3: Predict the French translation of a given English sentence.

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

# Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pro

Figure 4: Predict the 'win probability' of a given move in a state.



# How to “solve” problems without ML?

- ▶ Image classification:

- ▶ Recruit a “bird expert” to teach you about different birds features (e.g., beak shape, feather color, typical environment)
- ▶ Recognize these features in a given image, and then come up with a best guess of the bird species



<https://eduassistpro.github.io>

- ▶ Ask movie suppliers to categorize movies
- ▶ Predict a high rating for any movie in a user's good low rating for all other movies

- ▶ Machine translation: ...

- ▶ Chess: ...

- ▶ Applied ML
  - ▶ Collect/prepare data, build/train models, analyze performance/errors, etc
- ▶ ML developer

# Assignment Project Exam Help

- ▶ <https://eduassistpro.github.io>

Note: These roles are not mutually exclusive!

# Add WeChat edu\_assist\_pr

## Mathematical and computational prerequisites

- ▶ Math

- ▶ Linear algebra, probability, multivariable calculus

- ▶ Understand and reason about the concepts (not just calculations)



Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

## Basic setting: supervised learning

- ▶ Training data: dataset comprised of labeled examples
  - ▶ Labeled example: a pair of the form (input, label)
  - ▶ Input: what you see before you make a prediction (a.k.a. context, side-information, features, etc.)
  - ▶ Label: output value (a.k.a. output, response, target, etc.)

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

# Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pro

Figure 5: Decision tre

```
1: if  $0.335 \cdot x_1 + 2.5 \cdot x_2 + \dots + 6.35 \cdot x_{10^6} > 4.3$  then  
2:   return spam  
3: else  
4:   return not spam  
5: end if
```

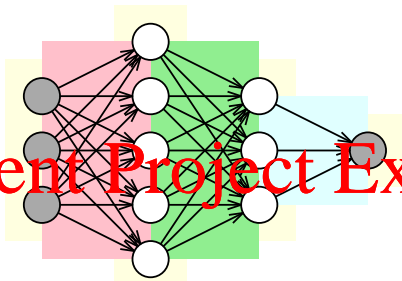
# Assignment Project Exam Help

Figure 6: Linear classifier ("Perceptron")

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

Assignment Project Exam Help



<https://eduassistpro.github.io>

Figure 7: Neural netwo

Add WeChat edu\_assist\_pr

# Types of prediction problems

- ▶ Binary classification
  - ▶ Given an email, is it spam or not?
  - ▶ (What's the probability that it is spam?)
- ▶ Multi-class classification

▶ <https://eduassistpro.github.io>

- ▶ (In absolute level? Log-scale?)
- ▶ Structured output prediction
  - ▶ Given a sentence, what is its grammatical structure?
  - ▶ (Or dependency tree?)
- ▶ ...



## Template of supervised learning pipeline

- ▶ Get data
- ▶ Determine representation of and predictive model for data
- ▶ Train the predictor (a.k.a. model fitting, parameter estimation)
- ▶ Evaluate predictor (test the “goodness of fit”)
- ▶

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

- ▶ What is the core prediction problem?

Assignment Project Exam Help

- ▶ What features (i.e., predictor variables) are available?

▶ <https://eduassistpro.github.io>

- ▶ Is there enough information (e.g., training data) to learn the relationship between the input and output?

Add WeChat edu\_assist\_pro

- ▶ What are the modeling assumptions?
- ▶ Is high-accuracy prediction a useful goal for the application?

- ▶ Form of the prediction function

## Assignment Project Exam Help



<https://eduassistpro.github.io>

- ▶ Choice of training data

Add WeChat edu\_assist\_pr

- ▶ Choice of learning algorithm
  - ▶ Choice of objective function and constraints

- ▶ Might not have the “right” data

# Assignment Project Exam Help

- ▶ Might not have the “right” model

- ▶ <https://eduassistpro.github.io>

- ▶ Add WeChat edu\_assist\_pr

- ▶ Data might be corrupted, noisy, ...

- ▶ Plug-in principle

# Assignment Project Exam Help



<https://eduassistpro.github.io>

- ▶ Linearity

Add WeChat edu\_assist\_pr

- ▶ Mathematical optimization

- ▶ Basic principles and methods of supervised machine learning
  1. Appetizer: nearest neighbor rules (a “non-parametric” method)
  2. Statistical model for prediction
  3. Regression
    - ▶ Why? Clean, simple, and illustrates important concepts

<https://eduassistpro.github.io>

- 6. Maybe one other topic if time permits ...
- ▶ This is not a course about how to use sklearn, etc.
- ▶ Also not about latest nonsense on arXiv
- ▶ Good stuff beyond COMS 4771:
  - ▶ COMS 4252, 4773: Mathematical theory of learning
  - ▶ COMS 4774: Unsupervised learning
  - ▶ COMS 4775: Causal inference
  - ▶ ...

- ▶ Professor Daniel Hsu
  - ▶ Okay to call me “Daniel”!
  - ▶ “Professor Hsu” also okay
  - ▶ “Professor Daniel” is a little weird

# Assignment Project Exam Help

<https://eduassistpro.github.io>

- <sup>A E</sup>  
▶ Bad at: making slides

# Add WeChat edu\_assist\_pr

## About you

- ▶ I assume you have fluency in
  - ▶ multivariable calculus,
  - ▶ linear algebra, and
  - ▶ elementary probability (no measure theory needed)



<https://eduassistpro.github.io>

brush up on Python, Numpy, etc.

- ▶ Let me know why you are interested in ML!
  - ▶ Part of HW 1.

Add WeChat edu\_assist\_pr



- ▶ Website: <https://www.cs.columbia.edu/~djhsu/ML>
  - ▶ Schedule for office hours/lectures/homework/quizzes/exam
  - ▶ Syllabus
  - ▶ Course format:

- ▶ <https://eduassistpro.github.io>

- ▶ Links for online office hours will be posted o
- ▶ Technology:
  - ▶ Piazza: communicate with course staff (
  - ▶ Courseworks: retrieve assignments, q
  - ▶ Gradescope: submit homework write-ups, code
  - ▶ Slack: discussion with fellow classmates
- ▶ Disability services:
  - ▶ Please make arrangements with disability services ASAP

## Academic rules of conduct

- ▶ See syllabus
- ▶ **Cheating:** don't do it
  - ▶ If unsure about something, ask!
  - ▶ Consequence is automatic fail



▶ <https://eduassistpro.github.io>



- ▶ No collaboration across groups
- ▶ No diffusion of responsibility
- ▶ No collaboration at all on quizzes or exams

## Reading assignments

- ▶ There are some required reading assignments (mostly from handouts posted on website)
- ▶ Unfortunately, most textbooks on ML are not appropriate for this course

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr