

# Assignment Project Exam Help

Machine learning lecture slides

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

Nearest neighbor classification

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

- ▶ Optical character recognition (OCR) example
- ▶ Nearest neighbor rule
- ▶ Error rate, test error rate
- ▶  $k$ -nearest neighbor rule
- ▶
- ▶
- ▶

# Assignment Project Exam Help

▶ <https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

## Example: OCR for digits

- ▶ Goal: Automatically label images of handwritten digits
- ▶ Possible labels are  $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$
- ▶ Start with a large collection of already-labeled images
- ▶  $D := \{(x_1, y_1), \dots, (x_n, y_n)\}$

ng

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

# Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

Figure 1: Some images of handwritten digits from

## Nearest neighbor (NN) classifier

- ▶ Nearest neighbor (NN) classifier  $NN_D$ :

- ▶ Represented using collection of labeled examples

$D := ((x_1, y_1), \dots, (x_n, y_n))$ , plus a snippet of code

- ▶ Input:  $x$

\_\_\_\_\_  $r$ )

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

## Naïve distance between images of handwritten digits (1)

- ▶ Treat (grayscale) images as vectors in Euclidean space  $\mathbb{R}^d$ 
  - ▶  $d = 28^2 = 784$
  - ▶ Generalizes physical 3-dimensional space
- ▶ Each point  $x = \underbrace{(x_1, \dots, x_d)}_d \in \underbrace{\mathbb{R}^d}_2$  is a vector of  $d$  real numbers

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

Figure 2: Grayscale pixel representation of an image of a handwritten “4”

## Naïve distance between images of handwritten digits (2)

- ▶ Why use this for images?

Assignment Project Exam Help

<https://eduassistpro.github.io>

- ▶ Why not use this for images?

Add WeChat edu\_assist\_pr



- ▶ What is the core prediction problem?

# Assignment Project Exam Help

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

- ▶ <https://eduassistpro.github.io>

- ▶ Is there enough information ("training data relationship between the features and lab

Add WeChat edu\_assist\_pr

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

- ▶ Error rate (on a collection of labeled examples  $S$ )
  - ▶ Fraction of labeled examples in  $S$  that have incorrect label prediction from  $\hat{f}$
  - ▶ Written as  $\text{err}(\hat{f}, S)$
  - ▶ (Often, the word “rate” is omitted)

Assignment Project Exam Help



<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

## Test error rate (1)

- ▶ Better evaluation: test error rate

- ▶ Train/test split,  $S \cap T = \emptyset$

- ▶  $S$  is training data,  $T$  is test data

- ▶ Classifier  $\hat{f}$  only based on  $S$

Assignment Project Exam Help

- ▶ <https://eduassistpro.github.io>

- ▶ What is the test error rate of uniformly  $r$

- ▶ What is the test error rate of a constant  $p$

Add WeChat edu\_assist\_pro

- ▶ Why is test error rate meaningful?

Assignment Project Exam Help

<https://eduassistpro.github.io>

- ▶ What are the drawbacks of evaluation via te

Add WeChat edu\_assist\_pr

# Assignment Project Exam Help

Figure <https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

# Assignment Project Exam Help

Figure <https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

# Assignment Project Exam Help

Figure <https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

# Assignment Project Exam Help

Figure <https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr



## More on the modeling assumptions

- ▶ Modeling assumption: Nearby images are more likely to have the same label than different labels.

Assignment Project Exam Help

- ▶ This is an assumption about the choice of distance function.
- ▶ In our OCR example, this is an assumption about the choice of features

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

- What are the kinds of errors made by  $NN_S$ ?

# Assignment Project Exam Help

Figure <https://eduassistpro.github.io>

## Add WeChat edu\_assist\_pr

Figure 8: Three nearest neighbors of the test example (8,2,2)

## Upgrade: $k$ -NN

►  $k$ -nearest neighbor ( $k$ -NN) classifier  $\text{NN}_{k,D}$

► Input:  $x$

- Find the  $k$  nearest neighbors of  $x$  in  $L$
- Return the plurality of the corresponding labels

►

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pro

## Typical effect of $k$

- ▶ Smaller  $k$ : smaller training error rate
- ▶ Larger  $k$ : higher training error rate, but predictions more “stable” due to voting
- ▶ On OCR data: lowest test error rate achieved at  $k = 3$

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

# Hyperparameter tuning

- ▶  $k$  is a hyperparameter of  $k$ -NN
- ▶ How to choose hyperparameters?
  - ▶ Bad idea: Choosing  $k$  that yields lowest training error rate (degenerate choice:  $k = 1$ )
  - ▶ Better idea: Simulate train/test split on the training data

▶ <https://eduassistpro.github.io>

$V_k := \text{err}(N$   
Add WeChat edu\_assist\_pr

- ▶ Let  $\hat{k}$  be the value of  $k$  for which
- ▶ Classifier to use is  $\text{NN}_{\hat{k}, S}$

## Upgrade: Distance functions (1)

- ▶ Specialize to input types
  - ▶ Edit distance for strings
  - ▶ Shape distance for images
  - ▶ Time warping distance for audio waveforms

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr

## Upgrade: Distance functions (2)

- ▶ Generic distances for vectors of real numbers

- ▶  $\ell_p$  distances

Assignment Project Exam Help

$$d_p(x, z) = \left( \sum_{j=1}^n |x_j - z_j|^p \right)^{1/p}.$$

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pro

## Upgrade: Distance functions (3)

- Distance functions for images of handwritten digits

	distance	$l_2$	tangent	shape
test error rate	0.0309	0.0283	0.0110	0.0065

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr



# Features

- ▶ When using numerical features (arranged in a vector from  $\mathbb{R}^d$ ):
  - ▶ Scale of features matters
  - ▶ Noisy features can ruin NN
  - ▶ E.g., consider what happens in OCR example if you have another 10000 additional features that are pure “noise”

the

<https://eduassistpro.github.io>

- ▶ “Curse of dimension”
  - ▶ Weird effects in  $\mathbb{R}^d$  for large  $d$
  - ▶ Can find  $2^{\Omega(d)}$  points that are approximately equidistant

- ▶ Brute force search:  $\Theta(dn)$  time for each prediction (using Euclidean distance in  $\mathbb{R}^d$ )
- ▶ Clever data structures: “improve” to  $2^d \log(n)$  time
- ▶ Approximate nearest neighbors: sub-linear time to get

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu\_assist\_pr