

Lec 1

Thursday, August 29, 2019 10:57

Supervised Learning

Given examples $x_1, y_1, \dots, x_n, y_n$

Want to learn a prediction rule f

s.t. for any new unseen example x, y
we have

$$y \approx f(x)$$

used loosely

can have different meanings

I.e., given a pair x, y but where

y is hidden, we need to guess
its value

Can be hidden b/c:

- unknowable
- will only realize in future
- need an expert to tell us what it is

Our first classification algorithm: KNN

Given a new query x :

- Find the k "closest" x -values among the n examples x_1, \dots, x_n
- Note their indices $i_1, \dots, i_k \in [n] = \{1, \dots, n\}$
- Classify as the majority of the corresponding labels:

$$\hat{y} = f(x) = \text{mode}(\{y_{i_1}, \dots, y_{i_k}\})$$

(most common value)

(break ties at random)

What does "closest" mean?

Need a distance measure