

DS2500
3/21 - Fri '21

Admin

- HW due 9pm
- HWS at, due 3/28
 - ↳ * 1st HW!
- Lab 5 Monday

- exam #2 4/4
(practice at 3/25)
- project due 4/5

- lecture 3/28
3:25 only
Zooming

Agenda

1. making predictions
2. evaluating supervised learning
3. Python

→ starter code on website

0. KNN Classifier

What is...?

- training/testing

- features/labels

- K - # neighbors to compare to

~ training - known values, used to fit the model

~ features - numeric attributes of every object

testing - predictions

label - category

1. making predictions

↳ supervised learning

KNN: look at features, predict label



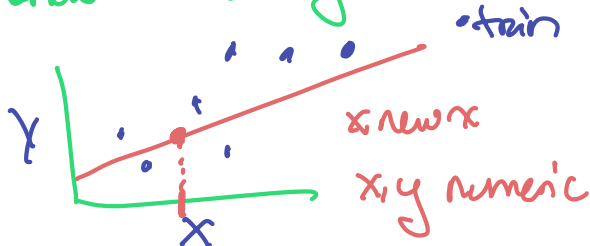
trues: k nearest for one test obj

today: k nearest for all test obj, make list of predictions

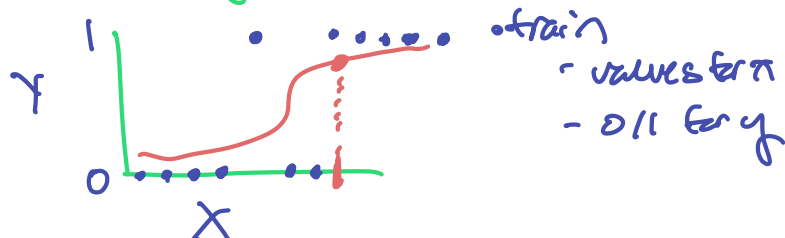
Supervised Learning algo #3

↳ logistic regression

we know: linear regression



new: logistic regr.



logistic regr is probabilistic:

$\geq .5$ yes / label

cat

$< .5$ no / not applied

not cat

any supervised learning...

1 - training

2 - fit

3 - testing / prediction

$$z = w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b$$

sigmoid function - squishes z to 0-1

In code (5 mins)

- make all preds in KNN
- actual, predicted on testing data

$x_{train}.csr$, $y_{train}.csr$

x_{train} features \rightarrow training
 y_{train} labels

x_{test} features \rightarrow testing
 y_{test} labels

\hookrightarrow ignored until evaluation!

2. Evaluating Supervised Learning

\hookrightarrow KNN evaluation: what makes best value of k ?

eval = computations, also need $\frac{0}{2}$

any supervised: some comparison of actual vs. predicted labels

- hide actual labels (y_{test})
- compute predictions
- see how well we did!

accuracy

precision

recall

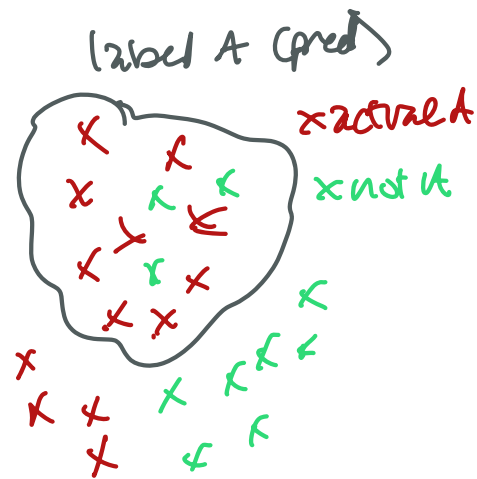
\rightarrow F1 score

accuracy ~ % of predictions that are correct

precision ~ % of things predicted label A are actually label A $9/12$

recall ~ % of actual label A's that I predicted as label A $9/13$

F1 ~ $\frac{\text{prec} * \text{recall}}{(\text{prec} + \text{recall})} * 2$



	name	friendliness	intelligence	label	predicted
0	Sven	8	3	dog	dog
1	Asteroid	3	7	cat	cat
2	Zeke	9	2	dog	dog
3	Melvin	5	4	dog	cat
4	Remy	6	6	dog	dog
5	Lady	5	7	dog	cat
6	Cam	9	3	dog	dog
7	Tito's	3	7	cat	cat
8	Bailey	3	7	cat	cat

accuracy
prec, recall
dogs

acc: 7/9 78%

prec: 4/4
100%

recall: 4/6
67%

accuracy
prec, recall
cats

acc: 78%

prec: 3/5
60%

recall: 3/3
100%

acc: 78% pretty good!

But... stupid model, everything is a dog 6/9! also pretty good

Confusion matrix predicted

		cat dog	
actual	cat	3	0
	dog	2	4

TP	FN
FP	TN