First, a true story, from
Greenwich Connecticut, 2007

First, a true story, from Greenwich Connecticut, 2007

Financial markets were at all-time highs (this is before the Great Financial Crisis)

Mr V worked at a quant hedge fund as a trader of credit derivatives.

Mr V was paid to build financial models, convince the hedge fund's owner that they were good models, and then trade them with the HFs' money

First, a true story, from Greenwich Connecticut, 2007

Financial markets were at all-time highs
(this is before the Great Financial Crisis)

Mr V worked at a quant hedge fund as a trader of credit derivatives.

Err..how?

Mr V was paid to build financial models, convince the hedge fund's owner that they were good models, and then trade them with the HFs' money

Using Backtests of course!

Using Backtests of course!

A backtest "runs" the model on recent market data, and tells how it performed.

Easy as Pie!!

Err..wasn't the model also built using recent market data?

Err..Yes..

Its really not an exaggeration that Overfitting ML models directly contributed to causing the GFC.

OVERFITTING IS THE BUGBEAR OF MACHINE LEARNING

SO WHAT IS OVERFITTING? AND WHY IS IT SUCH A PROBLEM?

CROSS VALIDATION

REGULARIZATION

ENSEMBLE LEARNING

SOME OF THE WAYS TO MITIGATE THIS PROBLEM

FRODO AND SAM ATE AT A RESTAURANT EVERY DAY LAST WEEK AND RATED IT ON EACH DAY

MONDAY	GOOD
TUESDAY	BAD
WEDNESDAY	GOOD
THURSDAY	GOOD
FRIDAY	GOOD
SATURDAY	BAD
SUNDAY	GOOD

FRODO AND SAM ATE AT A RESTAURANT EVERY DAY LAST WEEK AND RATED IT ON EACH DAY

MONDAY	GOOD
TUESDAY	BAD
WEDNESDAY	GOOD
THURSDAY	GOOD
FRIDAY	GOOD
SATURDAY	BAD
SUNDAY	GOOD

AT THE END OF THE WEEK,

FRODO SAYS

THE FOOD IS GOOD AT THIS RESTAURANT

SAM SAYS

THE FOOD IS GOOD AT THIS RESTAURANT ON ALL DAYS EXCEPT TUESDAYS AND SATURDAYS

WHICH ONE OF THEM IS RIGHT?

WHICH ONE OF THEM IS RIGHT?

HOW DO WE MEASURE THIS?

WE COULD CHECK EACH OF
THEIR STATEMENTS MODELS

AGAINST THE DATA WE ALREADY HAVE

TRAINING SET

WHICH ONE OF THEM IS RIGHT?

	TRAINING SET	FRODO'S MODEL	SAM'S MODEL
MONDAY	GOOD	GOOD	GOOD
TUESDAY	BAD	GOOD	BAD
WEDNESDAY	GOOD	GOOD	GOOD
THURSDAY	GOOD	GOOD	GOOD
FRIDAY	GOOD	GOOD	GOOD
SATURDAY	BAD	GOOD	BAD
SUNDAY	GOOD	GOOD	GOOD
		71%	100%

WE COULD CHECK EACH OF THEIR STATEMENTS
AGAINST THE DATA WE ALREADY HAVE

ACCURACY

WHICH ONE OF THEM IS RIGHT?

		FRODO'S MODEL	SAM'S MODEL
MONDAY	GOOD	GOOD	GOOD
TUESDAY	BAD	GOOD	BAD
WEDNESDAY	GOOD	71%	100%
THURSDAY	GOOD		
FRIDAY	GOOD	GOOD	GOOD
SATURDAY	BAD	GOOD	BAD
SUNDAY	GOOD	GOOD	GOOD
	ACCURACY	71%	100%

ON THE TRAINING SET, FRODO'S MODEL HAS 71% ACCURACY AND SAM'S MODEL HAS 100% ACCURACY

FROM THIS, IT SEEMS LIKE SAM'S MODEL IS BETTER.

SAM AND FRODO GO BACK TO THE RESTAURANT NEXT WEEK

WHICH ONE OF THEM IS RIGHT?

			FRODO'S MODEL	SAM'S MODEL
WEEK 1	MONDAY	GOOD	GOOD	GOOD
	TUESDAY	BAD	GOOD	BAD
	WEDNESDAY	GOOD	GOOD	GOOD
	THURSDAY	GOOD	71%	100%
	FRIDAY	GOOD	GOOD	GOOD
	SATURDAY	BAD	GOOD	BAD
	SUNDAY	GOOD	GOOD	GOOD
WEEK 2	MONDAY	GOOD	GOOD	GOOD
	TUESDAY	GOOD	GOOD	BAD
	WEDNESDAY	BAD	GOOD	GOOD
	THURSDAY	GOOD	71%	42%
	FRIDAY	GOOD	GOOD	GOOD
	SATURDAY	GOOD	GOOD	BAD
	SUNDAY	BAD	GOOD	GOOD

ON THE TRAINING SET, FRODO'S MODEL HAS 71% ACCURACY AND SAM'S MODEL HAS 100% ACCURACY

SAM AND FRODO GO BACK TO THE RESTAURANT NEXT WEEK

ON NEW DATA, FRODO'S MODEL HAS 71% ACCURACY AND SAM'S MODEL HAS 42% ACCURACY

WHICH ONE OF THEM IS RIGHT?

	FRODO'S MODEL	SAM'S MODEL
TRAINING SET	71%	100%
NEW/UNSEEN DATA	71%	42%

WHAT HAPPENED HERE?

FRODO'S MODEL IS THE BETTER MODEL

IT GENERALIZES WELL

FRODO'S MODEL PERFORMS WELL ON BOTH TRAINING AND NEW/UNSEEN DATA

WHAT HAPPENED HERE?

	FRODO'S MODEL	SAM'S MODEL
TRAINING SET	71%	100%
NEW/UNSEEN DATA	71%	42%
	THE FOOD IS GOOD AT THIS RESTAURANT	THE FOOD IS GOOD AT THIS RESTAURANT ON ALL DAYS EXCEPT TUESDAYS AND SATURDAYS

FRODO'S MODEL IS SIMPLER ("DUMBER", IN FACT), YET IT PERFORMS BETTER

SAM'S MODEL IS MORE COMPLEX, AND MORE ACCURATE ON THE TRAINING SET

YET, IT PERFORMS BADLY ON NEW DATA

IE, SAM'S MODEL DOES NOT GENERALIZE WELL

THE FOOD IS GOOD AT THIS RESTAURANT ON ALL DAYS EXCEPT TUESDAYS AND SATURDAYS

SAM'S MODEL PICKS UP ON A RELATIONSHIP BETWEEN THE WEEKDAY AND THE QUALITY OF FOOD

THIS RELATIONSHIP HOWEVER, IS SPECIFIC TO THE TRAINING SET, AND NOT TRUE IN GENERAL

SAM'S MODEL IS A PERFECT EXAMPLE OF

OVERFITTING

OVERFITTING OCCURS WHEN A MODEL PICKS UP ON RANDOM PHENOMENA OR NOISE PRESENT IN THE TRAINING SET INSTEAD OF THE UNDERLYING RELATIONSHIP BETWEEN THE INPUT AND OUTPUT

OVERFITTING

BUT WHY IS OVERFITTING SUCH A COMMON PROBLEM?

THE TRAINING SET IS ONLY PART OF A MUCH LARGER SET

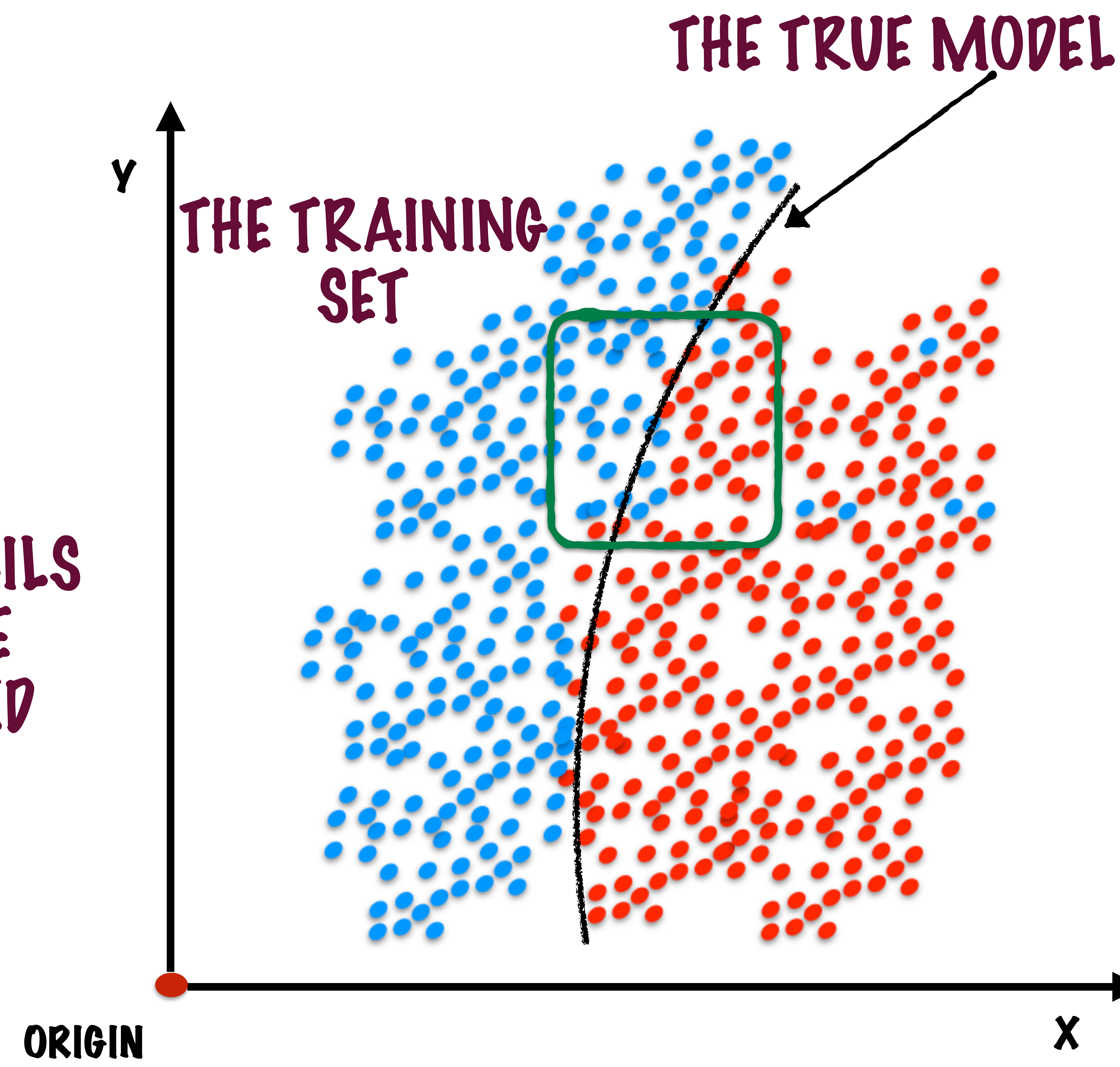
WE ARE TRYING TO FIND A MODEL, THAT DESCRIBES THIS MUCH LARGER SET

IT'S LIKE TRYING TO DESCRIBE PHOTOGRAPH, BUT YOU ARE ONLY SHOWN A SMALL, ZOOMED IN PORTION OF THE PHOTOGRAPH

OVERFITTING

YOU WANT TO CLASSIFY EMAILS AS SPAM OR HAM

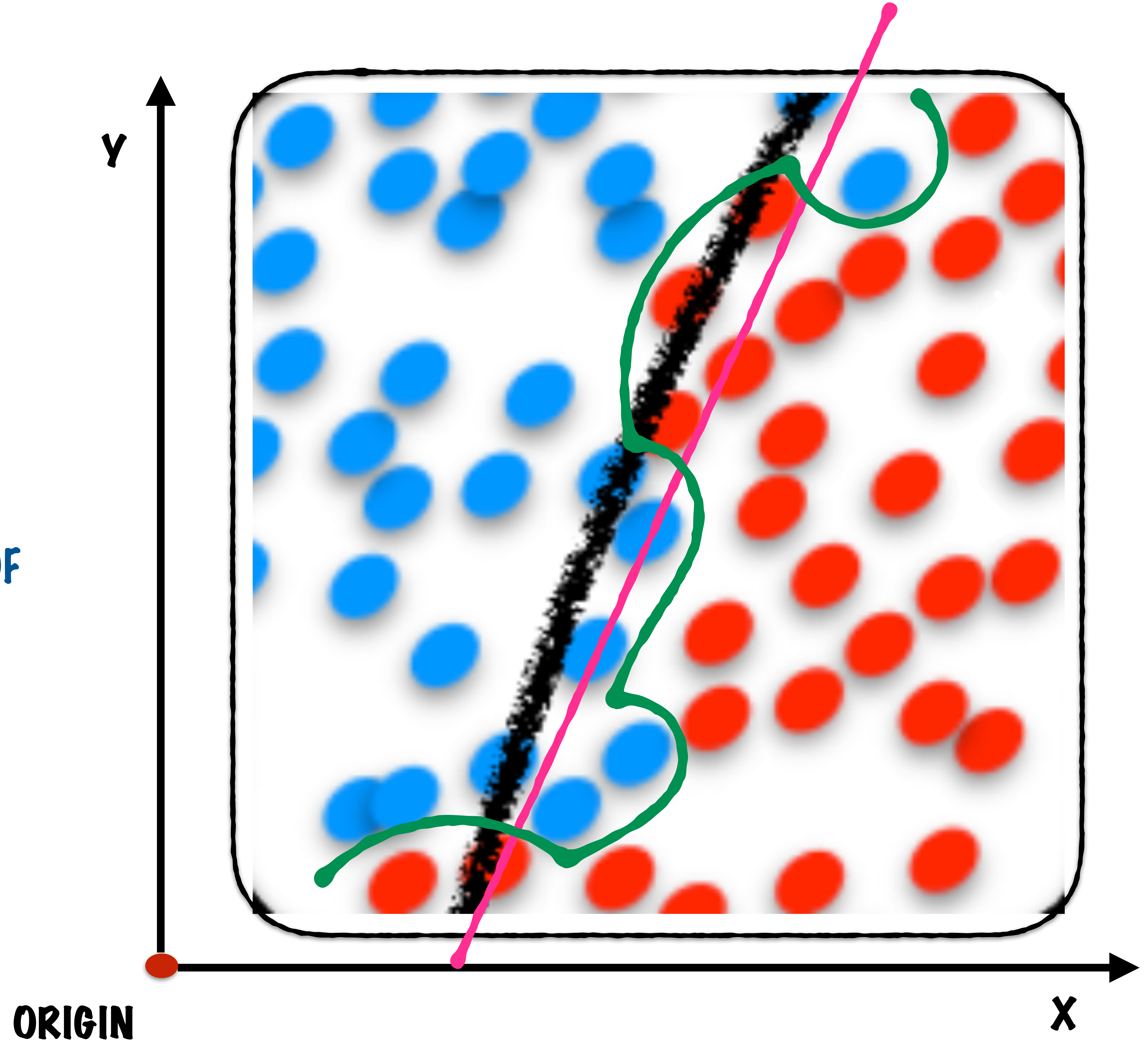
THESE ARE ALL THE EMAILS IN ALL INBOXES IN THE WORLD (BOTH PAST AND FUTURE)



OVERFITTING

1. A SIMPLE LINEAR MODEL

2. AN OVERFITTED MODEL
(USUALLY A POLYNOMIAL OF EXTREMELY HIGH ORDER)



BECAUSE THE TRAINING DATA IS ONLY A PART OF THE PICTURE

WE CAN'T TELL FOR SURE WHAT IS RELEVANT AND WHAT'S NOT

OVERFITTING

IS A PRETTY DIFFICULT PROBLEM TO SOLVE

BY AVOIDING OVERFITTING, WE CAN END UP WITH THE OPPOSITE ERROR OF UNDERFITTING

THIS IS THE FAMOUS BIAS-VARIANCE TRADEOFF

OVERFITTING IS THE BUGBEAR OF MACHINE LEARNING

SO WHAT IS OVERFITTING? AND WHY IS IT SUCH A PROBLEM?

CROSS VALIDATION

REGULARIZATION

ENSEMBLE LEARNING

SOME OF THE WAYS TO MITIGATE THIS PROBLEM

OVERFITTING IS THE BUGBEAR OF MACHINE LEARNING

SO WHAT IS OVERFITTING? AND WHY IS IT SUCH A PROBLEM?

CROSS VALIDATION

REGULARIZATION

ENSEMBLE LEARNING

SOME OF THE WAYS TO MITIGATE THIS PROBLEM

CROSS VALIDATION

IS A TECHNIQUE FOR MODEL SELECTION

PERFORMING WELL ON TRAINING DATA IS NO GUARANTEE FOR A GOOD MODEL

IN ORDER TO TEST THE PERFORMANCE OF A MODEL, IT WOULD BE NICE IF WE CAN

A GOOD MODEL IS ONE THAT PERFORMS WELL ON DATA IT HAS NOT SEEN BEFORE → GET SOME DATA THAT WE MIGHT SEE IN THE FUTURE (SOME NEW DATA)

GET MULTIPLE TRAINING DATA SETS

A GOOD MODEL DOES NOT OVERFIT → WE CAN THEN FIND A MODEL THAT PERFORMS WELL ACROSS TRAINING DATA SETS, AND NOT JUST ON ONE TRAINING SET

CROSS VALIDATION IS A COMBINATION OF THESE TWO IDEAS

IN ORDER TO TEST THE PERFORMANCE OF A MODEL, IT WOULD BE NICE IF WE CAN

GET SOME DATA THAT WE MIGHT SEE IN THE FUTURE (SOME NEW DATA) → KEEP SOME DATA ASIDE FOR PERFORMANCE TESTING

GET MULTIPLE TRAINING DATA SETS

WE CAN THEN FIND A MODEL THAT PERFORMS WELL ACROSS TRAINING DATA SETS, AND NOT JUST ON ONE TRAINING SET → CREATE MULTIPLE TRAINING SETS - EACH ONE A SUBSET OF THE ORIGINAL TRAINING SET

THE BELOW TABLE REPRESENTS THE ENTIRE TRAINING DATA SET

1. DIVIDE THE TRAINING SET RANDOMLY INTO TWO EQUAL PARTS - D_0 AND D_1

2. USE D_0 TO TRAIN THE MODEL AND D_1 TO TEST THE PERFORMANCE

D_0 | D_1

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}

TRAINING | TEST

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}

3. THEN, USE D_1 TO TRAIN THE MODEL AND D_0 TO TEST THE PERFORMANCE

THE BEST MODEL IS THE ONE WITH BEST AVERAGE PERFORMANCE

THIS TECHNIQUE IS CALLED

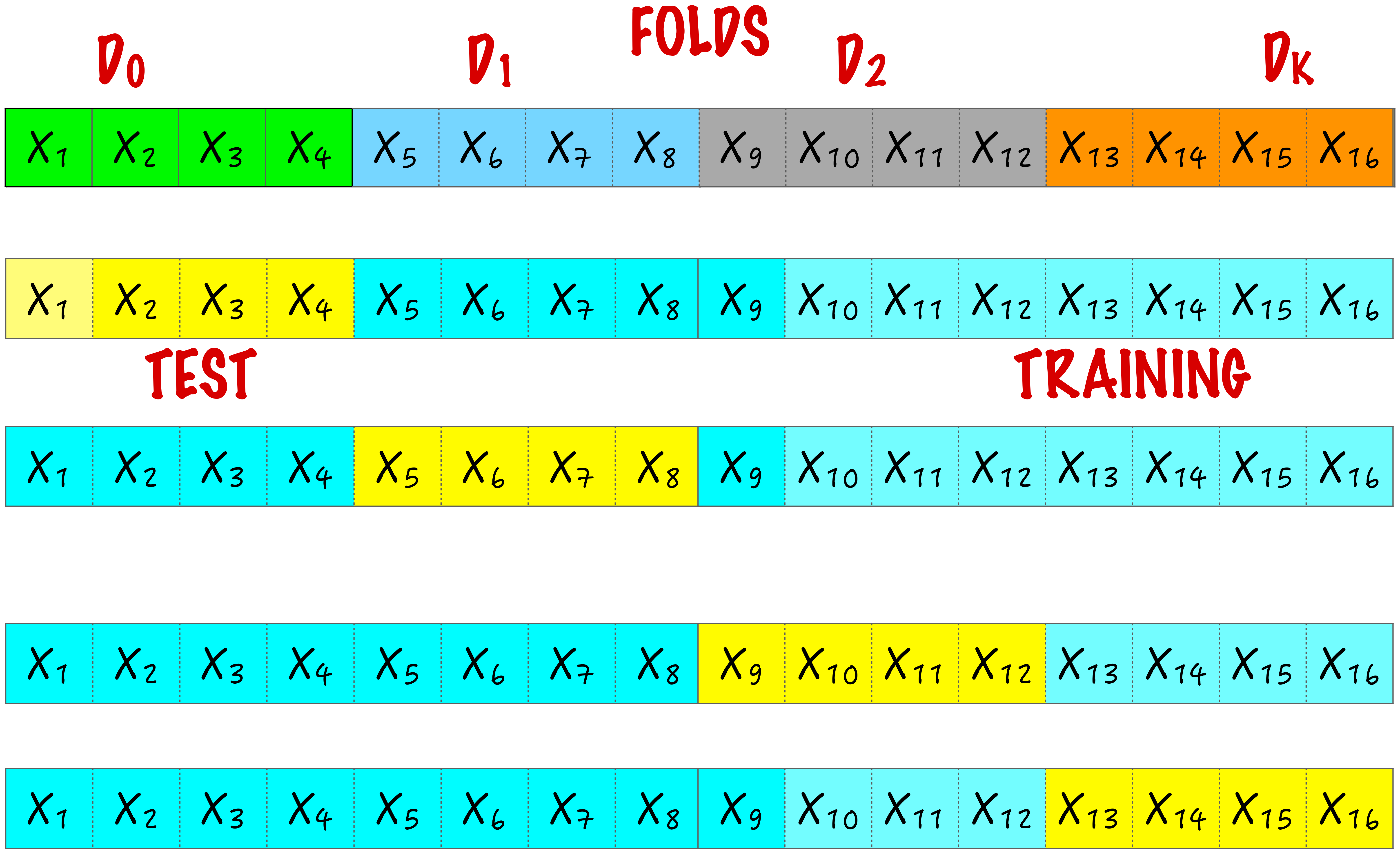
2-FOLD CROSS VALIDATION

K-FOLD CROSS VALIDATION

2-FOLD CROSS VALIDATION

1. DIVIDE THE TRAINING SET RANDOMLY INTO TWO EQUAL PARTS - D_0 AND D_1

1. DIVIDE THE TRAINING SET RANDOMLY INTO K EQUAL PARTS - $D_0, D_1, D_2, D_3, \ldots D_K$



2. USE D_0 TO TEST THE MODEL PERFORMANCE AND THE REST TO TRAIN THE DATA

3. USE D_1 TO TEST THE MODEL PERFORMANCE AND THE REST TO TRAIN THE DATA

4. CONTINUE UNTIL EACH OF THE PARTS HAS BEEN USED FOR TESTING EXACTLY ONCE

K-FOLD CROSS VALIDATION

A FEW NOTES

A COMMON CHOICE FOR K IS K=10

THE CHOICE OF K DEPENDS ON THE NUMBER OF SAMPLES (DATA POINTS) YOU HAVE IN THE TRAINING SET

CHOOSE K SUCH THAT THE NUMBER OF SAMPLES IN EACH FOLD IS NOT TOO SMALL

A FEW OTHER VARIANTS OF CROSS VALIDATION

LEAVE 1 OUT

LEAVE P OUT

MONTE CARLO

WHEN DO YOU USE CROSS VALIDATION?

1. TO CHOOSE BETWEEN DIFFERENT ALGORITHMS

SUPPORT VECTOR MACHINES VS K-NEAREST NEIGHBOURS

2. TO TUNE THE PARAMETERS OF THE ALGORITHM

THE VALUE OF K IN K-NEAREST NEIGHBOURS,
THE MAX DEPTH OF A DECISION TREE

3. TO IDENTIFY THE FEATURES THAT ARE RELEVANT

IF YOU HAVE 20 FEATURES, SHOULD YOU USE ALL
OF THEM? OR A SUBSET?

OVERFITTING IS THE BUGBEAR OF MACHINE LEARNING

SO WHAT IS OVERFITTING? AND WHY IS IT SUCH A PROBLEM?

CROSS VALIDATION

REGULARIZATION

ENSEMBLE LEARNING

SOME OF THE WAYS TO MITIGATE THIS PROBLEM

OVERFITTING IS THE BUGBEAR OF MACHINE LEARNING

SO WHAT IS OVERFITTING? AND WHY IS IT SUCH A PROBLEM?

CROSS VALIDATION

REGULARIZATION

ENSEMBLE LEARNING

SOME OF THE WAYS TO MITIGATE THIS PROBLEM

REGULARIZATION

PENALIZES MODELS WHICH ARE TOO COMPLEX

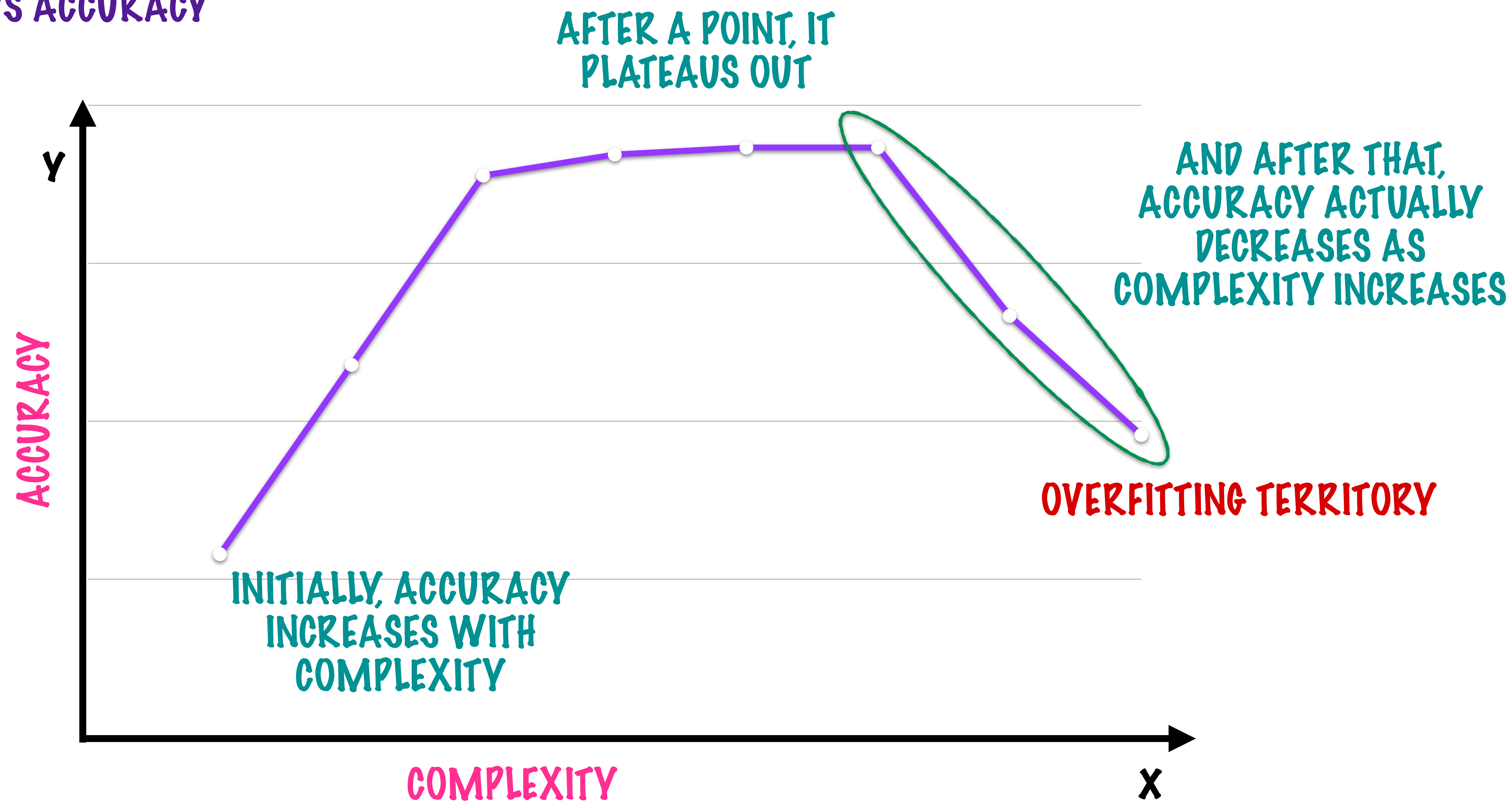
OVERFITTING OCCURS BECAUSE THE MODEL HAS BECOME NEEDLESSLY COMPLEX

EXAMPLES OF COMPLEXITY MEASURES

(THE NUMBER OF BRANCHES IN A DECISION TREE (OR) THE ORDER OF THE POLYNOMIAL USED TO REPRESENT A CURVE)

LET'S SAY YOU PLOTTED COMPLEXITY OF A MODEL VS ACCURACY

LET'S SAY YOU PLOTTED COMPLEXITY OF A MODEL VS ACCURACY



REGULARIZATION

PENALIZES MODELS WHICH ARE TOO COMPLEX

FINDING A MODEL USUALLY INVOLVES MINIMIZING AN ERROR FUNCTION

FOR EXAMPLE, THE ERROR FUNCTION COULD BE THE SUM OF SQUARES OF DISTANCES BETWEEN THE PREDICTED POINTS AND THE ACTUAL POINTS IN THE TRAINING SET

LET THE ERROR FUNCTION BE $E(f)$ FOR A MODEL f

LET THE ERROR FUNCTION BE $E(f)$ FOR A MODEL f

A REGULARIZATION TERM IS ADDED TO THIS FUNCTION

$$E'(f) = E(f) + \lambda R(f)$$

NEW ERROR FUNCTION THAT NEEDS TO BE MINIMIZED

A PARAMETER THAT CONTROLS THE IMPORTANCE OF THE REGULARIZATION TERM

REGULARIZATION TERM THAT INCREASES WITH COMPLEXITY

LET THE ERROR FUNCTION BE $E(f)$ FOR A MODEL f

A REGULARIZATION TERM IS ADDED TO THIS FUNCTION

$$\underbrace{E'(f)}_{\text{NEW ERROR FUNCTION THAT NEEDS TO BE MINIMIZED}} = E(f) + \underbrace{\lambda}_{\text{A PARAMETER THAT CONTROLS THE IMPORTANCE OF THE REGULARIZATION TERM}} \underbrace{R(f)}_{\text{REGULARIZATION TERM THAT INCREASES WITH COMPLEXITY}}$$

WE GET A MODEL THAT GIVES LOW ERROR ON THE TRAINING SET, WHILE KEEPING THE COMPLEXITY LOW AS WELL

ONE WELL-KNOWN EXAMPLE OF REGULARIZATION IS **ADJUSTED R-SQUARED**

IN LINEAR REGRESSION

R-SQUARED MEASURES HOW CLOSE THE DATA IN THE TRAINING SET IS TO THE LINE THAT'S FITTED

ADJUSTED R-SQUARED HAS BEEN ADJUSTED FOR THE NUMBER OF INDEPENDENT VARIABLES THAT HAVE BEEN USED IN THE MODEL

OVERFITTING IS THE BUGBEAR OF MACHINE LEARNING

SO WHAT IS OVERFITTING? AND WHY IS IT SUCH A PROBLEM?

CROSS VALIDATION

REGULARIZATION

SOME OF THE WAYS TO MITIGATE THIS PROBLEM

ENSEMBLE LEARNING

OVERFITTING IS THE BUGBEAR OF MACHINE LEARNING

SO WHAT IS OVERFITTING? AND WHY IS IT SUCH A PROBLEM?

CROSS VALIDATION

REGULARIZATION

SOME OF THE WAYS TO MITIGATE THIS PROBLEM

ENSEMBLE LEARNING

ENSEMBLE LEARNING

INVOLVES THE USE OF MULTIPLE LEARNERS AND COMBINING THEIR RESULTS

IN 2006, NETFLIX HELD AN OPEN COMPETITION FOR A MACHINE LEARNING ALGORITHM TO PREDICT A USER'S RATING OF A MOVIE

THE GRAND PRIZE WAS A COOL MILLION !

THE COMPETITION WENT ON FOR 3 YEARS, BEFORE A GRAND PRIZE WINNER WAS DECLARED

AN INTERESTING THING HAPPENED DURING THIS TIME...

THE CONTESTANTS FOUND THAT, INSTEAD OF USING 1 SINGLE MODEL, COMBINING MULTIPLE MODELS WORKED BETTER

TEAMS STARTED MERGING INTO LARGER TEAMS, THEY WOULD COMBINE THEIR MODELS TO DO BETTER

IN THE END, THE GRAND PRIZE WINNER (AND A VERY CLOSE RUNNER UP) WERE BOTH ENSEMBLES OF MORE THAN A 100 LEARNERS EACH..

AND COMBINING THEM IMPROVED THE RESULTS EVEN FURTHER!

THE IDEA OF ENSEMBLE LEARNING IS SIMPLE..

THE IDEA OF ENSEMBLE LEARNING IS SIMPLE..

MODELS TEND TO OVERFIT

IF YOU TRAIN MULTIPLE MODELS

THE OVERFITTING COMPONENTS OF EACH OF THE MODELS WOULD BE DIFFERENT

WHEN YOU COMBINE THESE MODELS

THE OVERFITTING COMPONENTS OF THE MODELS WOULD CANCEL EACH OTHER OUT

AND YOU ARE LEFT WITH THE COMPONENTS THAT REALLY DESCRIBE YOUR DATA

LET'S TAKE AN EXAMPLE

CLASSIFY A TWEET AS POSITIVE OR NEGATIVE SENTIMENT
(THIS IS A CLASSIFICATION PROBLEM)

METHOD 1. CHOOSE 1 TECHNIQUE

NAIVE BAYES (OR) SUPPORT VECTOR MACHINES (OR) NEURAL NETWORKS

METHOD 2. USE AN ENSEMBLE

NAIVE BAYES (AND) SUPPORT VECTOR MACHINES (AND) NEURAL NETWORKS

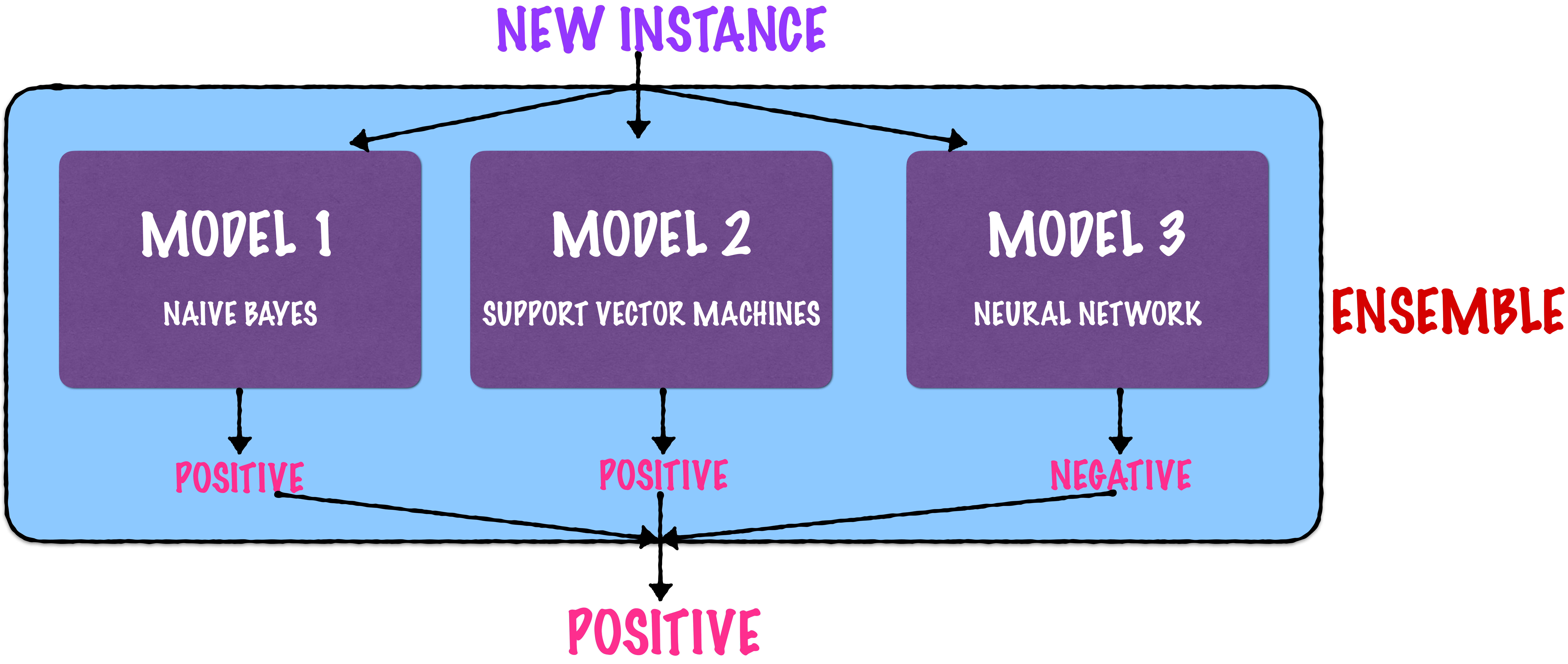
METHOD 2. USE AN ENSEMBLE

NAIVE BAYES (AND) SUPPORT VECTOR MACHINES (AND) NEURAL NETWORKS

1. TAKE THE TRAINING SET AND TRAIN EACH OF THE ABOVE CLASSIFIERS ON IT

2. WHEN A NEW INSTANCE (TWEET) COMES IN, GET THE PREDICTIONS FROM EACH OF THE MODELS

3. TAKE THE MAJORITY VOTE OF THE MODELS AND THAT WILL BE THE FINAL PREDICTION



A MACHINE LEARNING ENSEMBLE IS A COLLECTION OF MODELS

THE MODELS IN THE ENSEMBLE CAN BE

- BASED ON DIFFERENT TECHNIQUES
  - A COLLECTION WITH 1 SVM, 1 DECISION TREE, 1 NAIVE BAYES, 1 KNN
- TRAINED ON DIFFERENT TRAINING SETS
  - A COLLECTION OF SVMS, EACH TRAINED ON A DIFFERENT TRAINING SET
- USING DIFFERENT FEATURES
  - A COLLECTION OF DECISION TREES, EACH GIVEN A DIFFERENT SET OF FEATURES
- USING DIFFERENT VALUES OF PARAMETERS
  - A COLLECTION OF K-NEAREST NEIGHBOURS, EACH WITH A DIFFERENT VALUE OF K

A MACHINE LEARNING ENSEMBLE IS A COLLECTION OF MODELS

THE MODELS IN THE ENSEMBLE CAN BE

BASED ON DIFFERENT TECHNIQUES

TRAINED ON DIFFERENT TRAINING SETS

USING DIFFERENT FEATURES

USING DIFFERENT VALUES OF PARAMETERS

AN ENSEMBLE LEARNER COMBINES THE RESULTS FROM INDIVIDUAL MODELS

THE FINAL RESULT CAN BE

A MAJORITY VOTE OF THE INDIVIDUAL MODELS

AVERAGE OF THE RESULT FROM INDIVIDUAL MODELS

A WEIGHTED FUNCTION OF THE RESULT FROM INDIVIDUAL MODELS

A MACHINE LEARNING ENSEMBLE IS A COLLECTION OF MODELS

THE MODELS IN THE ENSEMBLE CAN BE

BASED ON DIFFERENT TECHNIQUES

TRAINED ON DIFFERENT TRAINING SETS

USING DIFFERENT FEATURES

USING DIFFERENT VALUES OF PARAMETERS

BAGGING

BOOSTING

ARE SPECIAL ENSEMBLE LEARNING TECHNIQUES

THEY INVOLVE CREATING MULTIPLE TRAINING SETS FROM THE MAIN TRAINING SET

AN ENSEMBLE LEARNER COMBINES THE RESULTS FROM INDIVIDUAL MODELS

THE FINAL RESULT CAN BE

A MAJORITY VOTE OF THE INDIVIDUAL MODELS

AVERAGE OF THE RESULT FROM INDIVIDUAL MODELS

A WEIGHTED FUNCTION OF THE RESULT FROM INDIVIDUAL MODELS

ONE OPTION IS TO WEIGHT EACH MODEL WITH IT'S ACCURACY

THIS MIGHT WORK, BUT IT ALSO RISKS OVERFITTING

THIS WEIGHTED FUNCTION CAN ALSO BE "LEARNED"

STACKING

INVOLVES TRAINING A LEARNER TO COMBINE THE RESULT FROM INDIVIDUAL MODELS

BAGGING

(BOOTSTRAP-AGGREGATING)

IS AN ENSEMBLE LEARNING TECHNIQUE THAT WAS DEVELOPED FOR CLASSIFICATION PROBLEMS

EACH MODEL IN THE ENSEMBLE IS TRAINED ON A DIFFERENT TRAINING SET

THESE TRAINING SETS ARE RANDOMLY GENERATED FROM THE ORIGINAL TRAINING SET

FOR THE FINAL RESULT, EACH MODEL IS GIVEN AN EQUAL WEIGHT AND A MAJORITY VOTE IS TAKEN

BAGGING
(BOOTSTRAP-AGGREGATING)

THESE TRAINING SETS ARE RANDOMLY GENERATED FROM THE ORIGINAL TRAINING SET

THE TRAINING SETS ARE GENERATED USING A STATISTICAL TECHNIQUE KNOWN AS **BOOTSTRAP SAMPLING**

SAMPLING IS THE PROCESS OF SELECTING INDIVIDUAL SAMPLES FROM A DATASET THAT ARE REPRESENTATIVE OF THE POPULATION

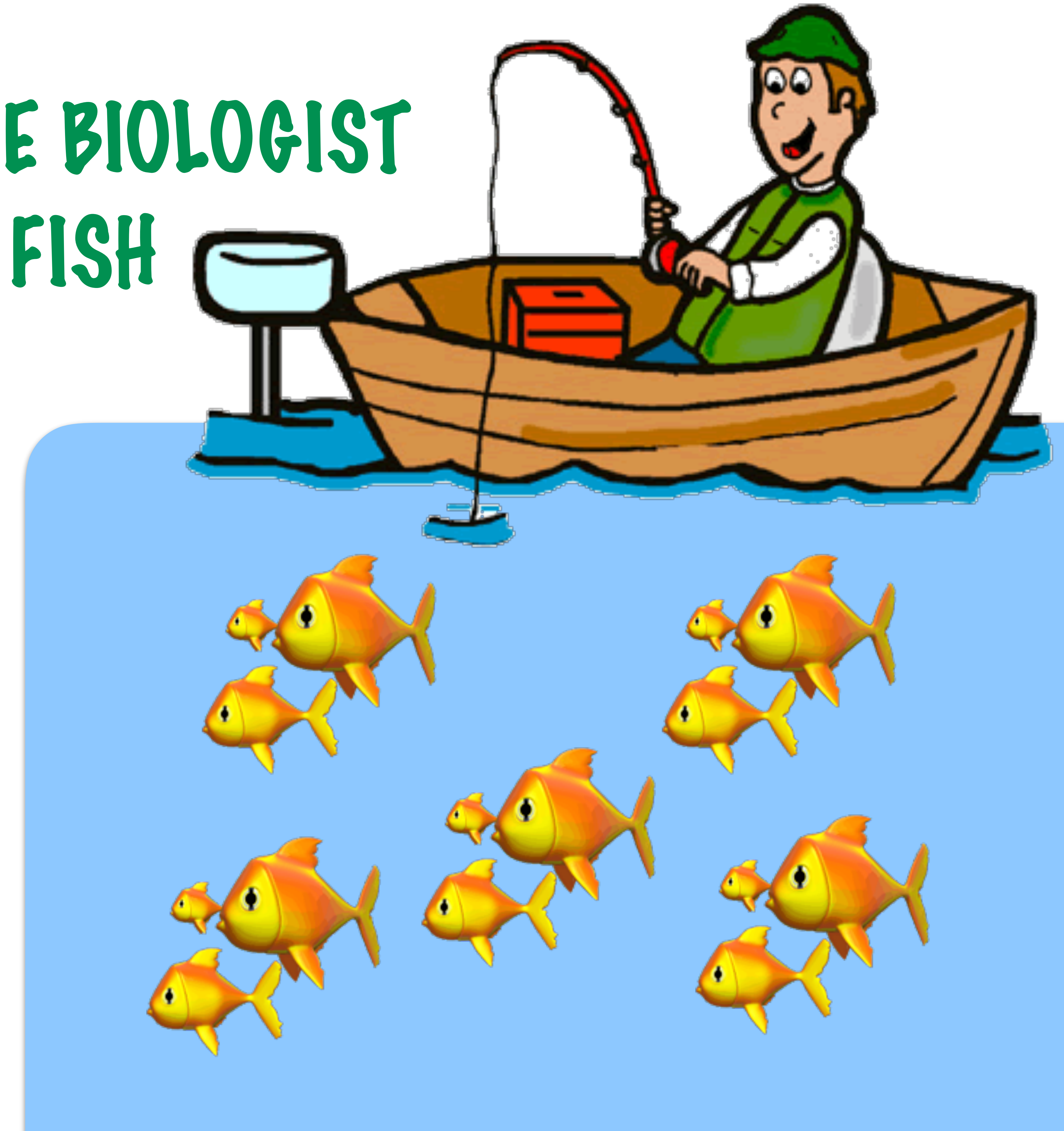
SAMPLING IS THE PROCESS OF SELECTING INDIVIDUAL SAMPLES FROM A DATASET THAT ARE REPRESENTATIVE OF THE POPULATION

DRAWING CONCLUSIONS ABOUT THE POPULATION FROM OBSERVATION OF THE SAMPLES IS CALLED

GENERALIZATION

SAMPLING IS A LITTLE BIT LIKE FISHING

YOU ARE A MARINE BIOLOGIST
YOU LEARN ABOUT FISH



YOU DO THIS BY

1. **CATCHING SOME FISH**
2. STUDYING THE CAUGHT FISH
3. DRAWING CONCLUSIONS ABOUT **ALL OF THE FISH**

POPULATION

SAMPLING IS THE PROCESS OF SELECTING INDIVIDUAL SAMPLES FROM A DATASET THAT ARE REPRESENTATIVE OF THE POPULATION

SAMPLING IS A LITTLE BIT LIKE FISHING

IF THE PROBABILITY OF CATCHING A PARTICULAR FISH, IS EXACTLY THE SAME AS ANY OTHER FISH

YOU ARE DOING

UNIFORM SAMPLING

IF YOU THROW THE FISH BACK IN AFTER YOU ARE DONE WITH IT

YOU ARE SAMPLING

WITH REPLACEMENT

SAMPLING IS THE PROCESS OF SELECTING INDIVIDUAL SAMPLES FROM A DATASET THAT ARE REPRESENTATIVE OF THE POPULATION

UNIFORM SAMPLING WITH REPLACEMENT

IS KNOWN AS **BOOTSTRAP SAMPLING**

BAGGING
(BOOTSTRAP-AGGREGATING)

IS AN ENSEMBLE LEARNING TECHNIQUE THAT USES BOOTSTRAP SAMPLING TO CREATE MULTIPLE TRAINING SETS

A MACHINE LEARNING ENSEMBLE IS A COLLECTION OF MODELS

THE MODELS IN THE ENSEMBLE CAN BE

BASED ON DIFFERENT TECHNIQUES

TRAINED ON DIFFERENT TRAINING SETS

USING DIFFERENT FEATURES

USING DIFFERENT VALUES OF PARAMETERS

AN ENSEMBLE LEARNER COMBINES THE RESULTS FROM INDIVIDUAL MODELS

THE FINAL RESULT CAN BE

A MAJORITY VOTE OF THE INDIVIDUAL MODELS

AVERAGE OF THE RESULT FROM INDIVIDUAL MODELS

A WEIGHTED FUNCTION OF THE RESULT FROM INDIVIDUAL MODELS

THIS WEIGHTED FUNCTION CAN ALSO BE "LEARNED"

BAGGING

BOOSTING

ARE SPECIAL ENSEMBLE LEARNING TECHNIQUES

THEY INVOLVE CREATING MULTIPLE TRAINING SETS FROM THE MAIN TRAINING SET

STACKING

INVOLVES TRAINING A LEARNER TO COMBINE THE RESULT FROM INDIVIDUAL MODELS

A MACHINE LEARNING ENSEMBLE IS A COLLECTION OF MODELS

THE MODELS IN THE ENSEMBLE CAN BE

BASED ON DIFFERENT TECHNIQUES

TRAINED ON DIFFERENT TRAINING SETS

USING DIFFERENT FEATURES

USING DIFFERENT VALUES OF PARAMETERS

AN ENSEMBLE LEARNER COMBINES THE RESULTS FROM INDIVIDUAL MODELS

THE FINAL RESULT CAN BE

A MAJORITY VOTE OF THE INDIVIDUAL MODELS

AVERAGE OF THE RESULT FROM INDIVIDUAL MODELS

A WEIGHTED FUNCTION OF THE RESULT FROM INDIVIDUAL MODELS

THIS WEIGHTED FUNCTION CAN ALSO BE "LEARNED"

BAGGING

BOOSTING

ARE SPECIAL ENSEMBLE LEARNING TECHNIQUES

THEY INVOLVE CREATING MULTIPLE TRAINING SETS FROM THE MAIN TRAINING SET

STACKING

INVOLVES TRAINING A LEARNER TO COMBINE THE RESULT FROM INDIVIDUAL MODELS

BOOSTING

IS AN ALGORITHM FOR ITERATIVELY ADDING LEARNERS TO THE ENSEMBLE

IN EACH ITERATION, THE TRAINING SET IS CHOSEN BY GIVING MORE WEIGHT TO THE MISCLASSIFIED SAMPLES

EACH LEARNER IN THE ENSEMBLE WILL PERFORM VERY POORLY BY ITSELF

WEAK LEARNERS

THE THEORY OF BOOSTING IS THAT AN ENSEMBLE OF WEAK LEARNERS CAN TOGETHER BE VERY STRONG

ADABOOST

IS THE MOST WELL KNOWN VARIANT OF A BOOSTING ALGORITHM

WHEN PREDICTING FOR A NEW INSTANCE, ADABOOST USES AN OPTIMALLY WEIGHTED VOTE OF THE ENSEMBLE LEARNERS

A MACHINE LEARNING ENSEMBLE IS A COLLECTION OF MODELS

THE MODELS IN THE ENSEMBLE CAN BE

BASED ON DIFFERENT TECHNIQUES

TRAINED ON DIFFERENT TRAINING SETS

USING DIFFERENT FEATURES

USING DIFFERENT VALUES OF PARAMETERS

AN ENSEMBLE LEARNER COMBINES THE RESULTS FROM INDIVIDUAL MODELS

THE FINAL RESULT CAN BE

A MAJORITY VOTE OF THE INDIVIDUAL MODELS

AVERAGE OF THE RESULT FROM INDIVIDUAL MODELS

A WEIGHTED FUNCTION OF THE RESULT FROM INDIVIDUAL MODELS

THIS WEIGHTED FUNCTION CAN ALSO BE "LEARNED"

BAGGING

BOOSTING

ARE SPECIAL ENSEMBLE LEARNING TECHNIQUES

THEY INVOLVE CREATING MULTIPLE TRAINING SETS FROM THE MAIN TRAINING SET

STACKING

INVOLVES TRAINING A LEARNER TO COMBINE THE RESULT FROM INDIVIDUAL MODELS

A MACHINE LEARNING ENSEMBLE IS A COLLECTION OF MODELS

THE MODELS IN THE ENSEMBLE CAN BE

BASED ON DIFFERENT TECHNIQUES

TRAINED ON DIFFERENT TRAINING SETS

USING DIFFERENT FEATURES

USING DIFFERENT VALUES OF PARAMETERS

AN ENSEMBLE LEARNER COMBINES THE RESULTS FROM INDIVIDUAL MODELS

THE FINAL RESULT CAN BE

A MAJORITY VOTE OF THE INDIVIDUAL MODELS

AVERAGE OF THE RESULT FROM INDIVIDUAL MODELS

A WEIGHTED FUNCTION OF THE RESULT FROM INDIVIDUAL MODELS

THIS WEIGHTED FUNCTION CAN ALSO BE "LEARNED"

BAGGING

BOOSTING

ARE SPECIAL ENSEMBLE LEARNING TECHNIQUES

THEY INVOLVE CREATING MULTIPLE TRAINING SETS FROM THE MAIN TRAINING SET

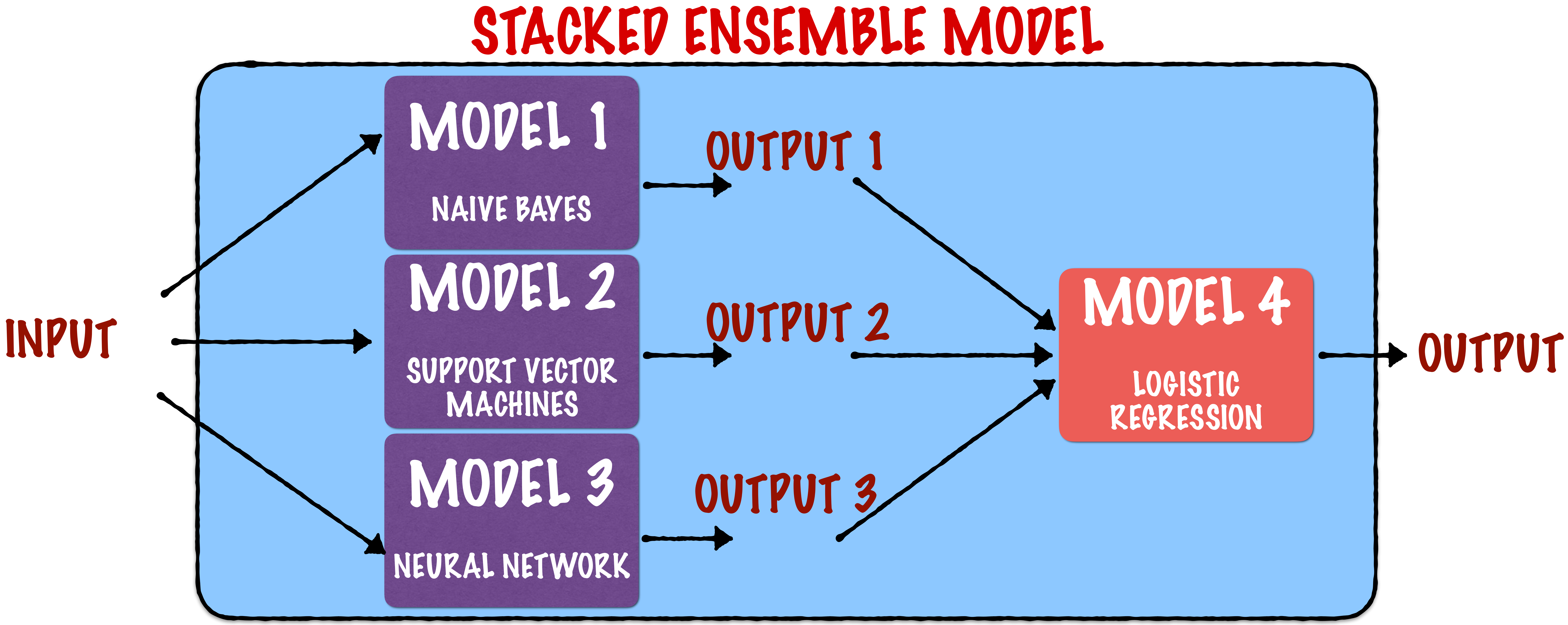
STACKING

INVOLVES TRAINING A LEARNER TO COMBINE THE RESULT FROM INDIVIDUAL MODELS

STACKING

(AKA BLENDING AKA STACKED GENERALIZATION)

INVOLVES USING A MACHINE LEARNING APPROACH TO COMBINE THE RESULTS OF THE ENSEMBLE MEMBERS



A MACHINE LEARNING ENSEMBLE IS A COLLECTION OF MODELS

THE MODELS IN THE ENSEMBLE CAN BE

BASED ON DIFFERENT TECHNIQUES

TRAINED ON DIFFERENT TRAINING SETS

USING DIFFERENT FEATURES

USING DIFFERENT VALUES OF PARAMETERS

AN ENSEMBLE LEARNER COMBINES THE RESULTS FROM INDIVIDUAL MODELS

THE FINAL RESULT CAN BE

A MAJORITY VOTE OF THE INDIVIDUAL MODELS

AVERAGE OF THE RESULT FROM INDIVIDUAL MODELS

A WEIGHTED FUNCTION OF THE RESULT FROM INDIVIDUAL MODELS

THIS WEIGHTED FUNCTION CAN ALSO BE "LEARNED"

BAGGING

BOOSTING

ARE SPECIAL ENSEMBLE LEARNING TECHNIQUES

THEY INVOLVE CREATING MULTIPLE TRAINING SETS FROM THE MAIN TRAINING SET

STACKING

INVOLVES TRAINING A LEARNER TO COMBINE THE RESULT FROM INDIVIDUAL MODELS

A MACHINE LEARNING ENSEMBLE IS A COLLECTION OF MODELS

THE MODELS IN THE ENSEMBLE CAN BE

BASED ON DIFFERENT TECHNIQUES

TRAINED ON DIFFERENT TRAINING SETS

USING DIFFERENT FEATURES

USING DIFFERENT VALUES OF PARAMETERS

AN ENSEMBLE LEARNER COMBINES THE RESULTS FROM INDIVIDUAL MODELS

THE FINAL RESULT CAN BE

A MAJORITY VOTE OF THE INDIVIDUAL MODELS

AVERAGE OF THE RESULT FROM INDIVIDUAL MODELS

A WEIGHTED FUNCTION OF THE RESULT FROM INDIVIDUAL MODELS

THIS WEIGHTED FUNCTION CAN ALSO BE "LEARNED"

BAGGING

BOOSTING

ARE SPECIAL ENSEMBLE LEARNING TECHNIQUES

THEY INVOLVE CREATING MULTIPLE TRAINING SETS FROM THE MAIN TRAINING SET

STACKING

INVOLVES TRAINING A LEARNER TO COMBINE THE RESULT FROM INDIVIDUAL MODELS

OVERFITTING IS THE BUGBEAR OF MACHINE LEARNING

SO WHAT IS OVERFITTING? AND WHY IS IT SUCH A PROBLEM?

CROSS VALIDATION

REGULARIZATION

SOME OF THE WAYS TO MITIGATE THIS PROBLEM

ENSEMBLE LEARNING

OVERFITTING

IS THE BUGBEAR OF MACHINE LEARNING

SO WHAT IS OVERFITTING? AND WHY IS IT SUCH A PROBLEM?

CROSS VALIDATION

REGULARIZATION

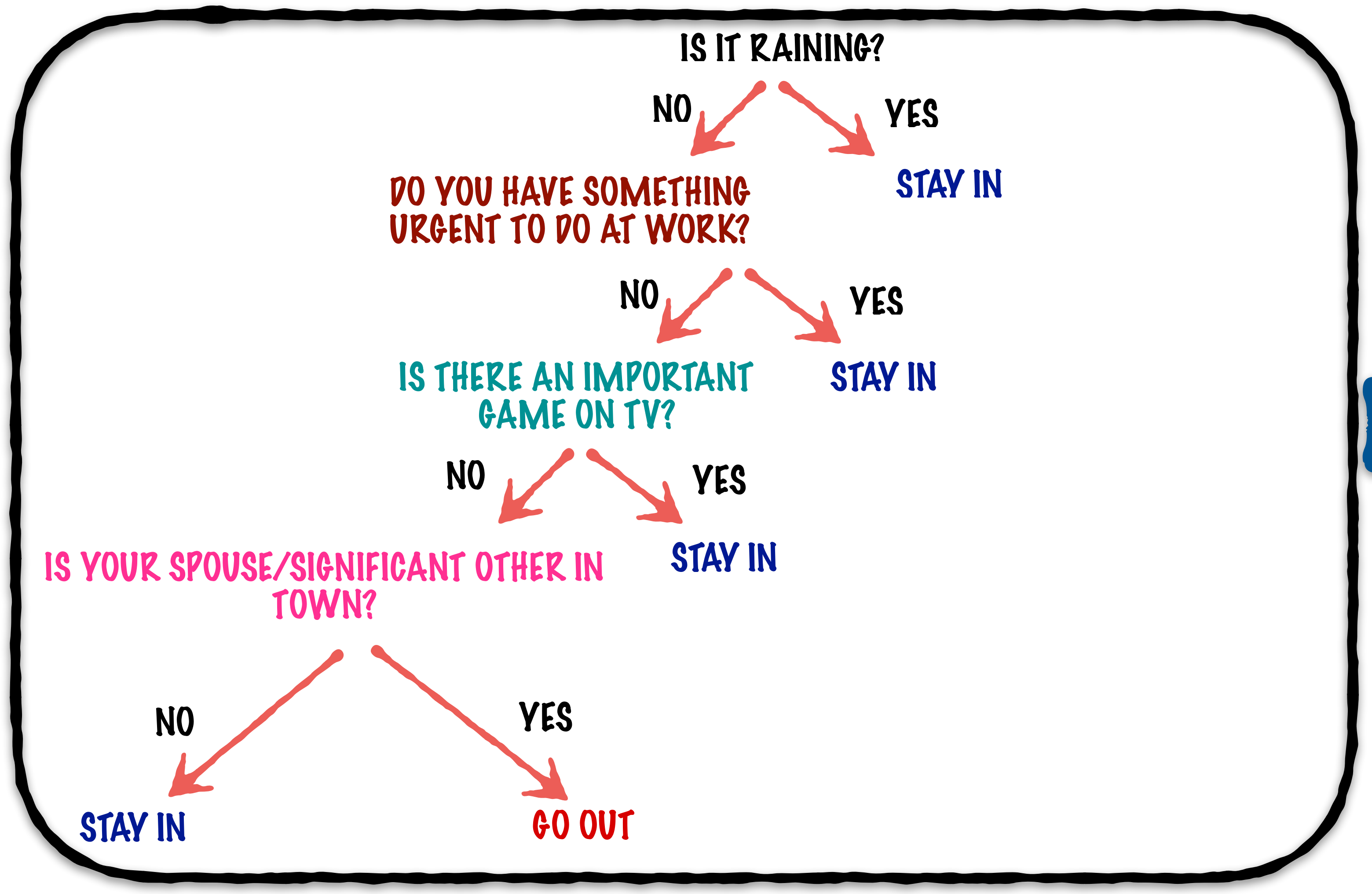
SOME OF THE WAYS TO MITIGATE THIS PROBLEM

ENSEMBLE LEARNING

A DECISION TREE CAN BE USED TO SOLVE MACHINE LEARNING PROBLEMS

A DECISION TREE PREDICTS THE OUTCOME GIVEN THE VALUES OF INPUT VARIABLES

INPUT VARIABLES/ PREDICTORS



OUTCOME/OUTPUT VARIABLES

DECISION TREES ARE VERY PRONE TO THE RISK OF OVERFITTING

ENSEMBLE LEARNING CAN MITIGATE THE RISK OF OVERFITTING

RECAP

ENSEMBLE LEARNING

RECAP

ENSEMBLE LEARNING

INVOLVES THE USE OF MULTIPLE LEARNERS AND COMBINING THEIR RESULTS

THE IDEA OF ENSEMBLE LEARNING IS SIMPLE..

MODELS TEND TO OVERFIT

IF YOU TRAIN MULTIPLE MODELS

THE OVERFITTING COMPONENTS OF EACH OF THE MODELS WOULD BE DIFFERENT

WHEN YOU COMBINE THESE MODELS

THE OVERFITTING COMPONENTS OF THE MODELS WOULD CANCEL EACH OTHER OUT

AND YOU ARE LEFT WITH THE COMPONENTS THAT REALLY DESCRIBE YOUR DATA

RECAP

A MACHINE LEARNING ENSEMBLE IS A COLLECTION OF MODELS

THE MODELS IN THE ENSEMBLE CAN BE

- BASED ON DIFFERENT TECHNIQUES
- TRAINED ON DIFFERENT TRAINING SETS
- USING DIFFERENT FEATURES
- USING DIFFERENT VALUES OF PARAMETERS

A MACHINE LEARNING ENSEMBLE IS A COLLECTION OF MODELS

THE MODELS IN THE ENSEMBLE CAN BE

BASED ON DIFFERENT TECHNIQUES

TRAINED ON DIFFERENT TRAINING SETS

USING DIFFERENT FEATURES

USING DIFFERENT VALUES OF PARAMETERS

A RANDOM FOREST IS AN ENSEMBLE OF DECISION TREES

EACH DECISION TREE IN THE ENSEMBLE IS

(BAGGING)
TRAINED ON DIFFERENT TRAINING SETS

USING DIFFERENT FEATURES
(A RANDOMLY SELECTED SUBSET OF FEATURES)

RECAP

BAGGING

(BOOTSTRAP-AGGREGATING)

RECAP

BAGGING

(BOOTSTRAP-AGGREGATING)

IS AN ENSEMBLE LEARNING TECHNIQUE THAT WAS DEVELOPED FOR CLASSIFICATION PROBLEMS

EACH MODEL IN THE ENSEMBLE IS TRAINED ON A DIFFERENT TRAINING SET

BOOTSTRAP SAMPLING

THESE TRAINING SETS ARE RANDOMLY GENERATED FROM THE ORIGINAL TRAINING SET

FOR THE FINAL RESULT, EACH MODEL IS GIVEN AN EQUAL WEIGHT AND A MAJORITY VOTE IS TAKEN

RECAP

BOOTSTRAP SAMPLING

SAMPLING IS THE PROCESS OF SELECTING INDIVIDUAL SAMPLES FROM A DATASET THAT ARE REPRESENTATIVE OF THE POPULATION

INVOLVES

UNIFORM SAMPLING

EACH SAMPLE HAS EQUAL PROBABILITY OF BEING CHOSEN IN THE TRAINING SET

WITH REPLACEMENT

ONCE A SAMPLE IS CHOSEN, IT IS ADDED BACK TO THE ORIGINAL SET, SO IT CAN BE CHOSEN AGAIN

RANDOM FORESTS

USE BOOTSTRAP SAMPLING TO CREATE MULTIPLE TRAINING SETS

EACH DECISION TREE IN THE ENSEMBLE IS TRAINED WITH A DIFFERENT ONE OF THESE TRAINING SETS

A MACHINE LEARNING ENSEMBLE IS A COLLECTION OF MODELS

THE MODELS IN THE ENSEMBLE CAN BE

BASED ON DIFFERENT TECHNIQUES

TRAINED ON DIFFERENT TRAINING SETS

USING DIFFERENT FEATURES

USING DIFFERENT VALUES OF PARAMETERS

A RANDOM FOREST IS AN ENSEMBLE OF DECISION TREES

EACH DECISION TREE IN THE ENSEMBLE IS

(BAGGING)
TRAINED ON DIFFERENT TRAINING SETS

USING DIFFERENT FEATURES
(A RANDOMLY SELECTED SUBSET OF FEATURES)

A MACHINE LEARNING ENSEMBLE IS A COLLECTION OF MODELS

THE MODELS IN THE ENSEMBLE CAN BE

BASED ON DIFFERENT TECHNIQUES

TRAINED ON DIFFERENT TRAINING SETS

USING DIFFERENT FEATURES

USING DIFFERENT VALUES OF PARAMETERS

A RANDOM FOREST IS AN ENSEMBLE OF DECISION TREES

EACH DECISION TREE IN THE ENSEMBLE IS

(BAGGING)

✓ TRAINED ON DIFFERENT TRAINING SETS

$\boxed{\text{USING DIFFERENT FEATURES (A RANDOMLY SELECTED SUBSET OF FEATURES)}}$

SAY YOU HAVE 7 POSSIBLE FEATURES TO USE TO PREDICT SURVIVAL ON THE TITANIC

GENDER

AGE

PORT OF EMBARKATION

SIBLINGS

PARENTS

PASSENGER CLASS

FARE

RANDOM SUBSPACE METHOD

EACH DECISION TREE IN THE RANDOM FOREST IS GIVEN A DIFFERENT SUBSET OF THESE 7 FEATURES TO LEARN FROM

THIS SUBSET IS RANDOMLY CHOSEN

A MACHINE LEARNING ENSEMBLE IS A COLLECTION OF MODELS

THE MODELS IN THE ENSEMBLE CAN BE

BASED ON DIFFERENT TECHNIQUES

TRAINED ON DIFFERENT TRAINING SETS

USING DIFFERENT FEATURES

USING DIFFERENT VALUES OF PARAMETERS

A RANDOM FOREST IS AN ENSEMBLE OF DECISION TREES

EACH DECISION TREE IN THE ENSEMBLE IS

✓ TRAINED ON DIFFERENT TRAINING SETS (BAGGING)

✓ USING DIFFERENT FEATURES (A RANDOMLY SELECTED SUBSET OF FEATURES)

A MACHINE LEARNING ENSEMBLE IS A COLLECTION OF MODELS

THE MODELS IN THE ENSEMBLE CAN BE

BASED ON DIFFERENT TECHNIQUES

TRAINED ON DIFFERENT TRAINING SETS

USING DIFFERENT FEATURES

USING DIFFERENT VALUES OF PARAMETERS

A RANDOM FOREST IS AN ENSEMBLE OF DECISION TREES

EACH DECISION TREE IN THE ENSEMBLE IS

- ✓ TRAINED ON DIFFERENT TRAINING SETS (BAGGING)
- ✓ USING DIFFERENT FEATURES (A RANDOMLY SELECTED SUBSET OF FEATURES)

RANDOM FOREST